



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> : <b>C12N 1/21, 5/10, 15/11, 15/12, 15/63, A61K 38/16, 38/17, C07K 14/00, 14/435, 16/00, G01N 33/50</b>		A1	(11) International Publication Number: <b>WO 00/06698</b>
			(43) International Publication Date: 10 February 2000 (10.02.00)
(21) International Application Number: PCT/US99/17130		437 West Side Drive, Gaithersburg, MD 20878 (US). LAFLEUR, David [US/US]; 3142 Quesada Street, N.W., Washington, DC 20015 (US). WEI, Ying-Fei [CN/US]; 242 Gravatt Drive, Berkeley, CA 94705 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). FLORENCE, Kimberly, A. [US/US]; 12805 Atlantic Avenue, Rockville, MD 20851 (US). YOUNG, Paul, E. [US/US]; 122 Beckwith Street, Gaithersburg, MD 20878 (US). BREWER, Laurie, A. [US/US]; Apartment 115, 410 Van Dyke Street, St. Paul, MN 55119-4321 (US). SOPPET, Daniel, R. [US/US]; 15050 Stillfield Place, Centreville, VA 22020 (US). ENDRESS, Gregory, A. [US/US]; 9729 Clagett Farm Drive, Potomac, MD 20854 (US). EBNER, Reinhard [DE/US]; 9906 Shelburne Terrace #316, Gaithersburg, MD 20878 (US). OLSEN, Henrik, S. [DK/US]; 182 Kendrick Place #24, Gaithersburg, MD 20878 (US). MUCENSKI, Michael [US/US]; 3265 Mandale Drive, Cincinnati, OH 45239 (US).	
(22) International Filing Date: 29 July 1999 (29.07.99)			
(30) Priority Data:			
60/094,657 30 July 1998 (30.07.98) US 60/095,486 5 August 1998 (05.08.98) US 60/095,455 6 August 1998 (06.08.98) US 60/095,454 6 August 1998 (06.08.98) US 60/096,319 12 August 1998 (12.08.98) US			
(71) Applicant (for all designated States except US): HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).			
(72) Inventors; and		(74) Agents: HOOVER, Kenley, K. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).	
(75) Inventors/Applicants (for US only): KOMATSOUKIS, George, A. [US/US]; 9518 Garwood Street, Silver Spring, MD 20901 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). DUAN, Roxanne [US/US]; 5515 Northfield Road, Bethesda, MD 20817 (US). MOORE, Paul, A. [GB/US]; 19005 Leatherbark Drive, Germantown, MD 20874 (US). SHI, Yanggu [CN/US]; Apartment 102,			
		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
		Published With international search report. With an indication in relation to deposited biological material furnished under Rule 13bis separately from the description.	
(54) Title: 98 HUMAN SECRETED PROTEINS			
(57) Abstract			
The present invention relates to novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		



## 98 Human Secreted Proteins

### *Field of the Invention*

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and  
5 their production.

### *Background of the Invention*

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or  
10 organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic  
15 reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

20 Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in  
25 secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins  
30 include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of

the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

5

### *Summary of the Invention*

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and  
10 polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

15

### *Detailed Description*

#### Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original  
20 environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

25 In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing  
30 to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

In specific embodiments, the polynucleotides of the invention are less than 300 kb, 200 kb, 100 kb, 50 kb, 15 kb, 10 kb, or 7.5 kb in length. In a further embodiment, polynucleotides of the invention comprise at least 15 contiguous nucleotides of the coding sequence, but do not comprise all or a portion of any intron.

- 5 In another embodiment, the nucleic acid comprising the coding sequence does not contain coding sequences of a genomic flanking gene (i.e., 5' or 3' to the gene in the genome).

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone  
10 deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a  
15 molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig analysis). A representative clone containing all or most of the sequence for SEQ ID  
20 NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard, Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of  
25 microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an  
30 overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's

solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions.

5 Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M

10  $\text{NaH}_2\text{PO}_4$ ; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the

15 inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above,

20 due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid

25 molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of

30 single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA

that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

(See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth  
5 Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting  
10 activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the  
15 present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

## 20 Polynucleotides and Polypeptides of the Invention

### FEATURES OF PROTEIN ENCODED BY GENE NO: 1

The translation product of this gene is a human glycoprotein-associated amino acid transporter (See, e.g., Genbank Accession No. emb|CAA10198.1| (AJ130718);  
25 all references available through this accession are hereby incorporated by reference herein). Amino acid transport across cellular membranes is mediated by multiple transporters with overlapping specificities. The transport system L, which mediates Na<sup>+</sup>-independent exchange of large neutral amino acids, consists of a novel amino acid permease-related protein (LAT1 or AmAT-L-lc) which for surface expression  
30 and function requires formation of disulfide-linked heterodimers with the glycosylated heavy chain of the h4F2/CD98 surface antigen. h4F2hc also associates with other mammalian light chains, e.g. y+LAT1 from mouse and human which are

approximately 48% identical with LAT1 and thus belong to the same family of glycoprotein-associated amino acid transporters.

The novel heterodimers form exchangers which mediate the cellular efflux of cationic amino acids and the Na<sup>+</sup>-dependent uptake of large neutral amino acids.

5 These transport characteristics and kinetic and pharmacological fingerprints identify them as y<sup>+</sup>L-type transport systems. mRNA encoding y<sup>+</sup>LAT1 is detectable in most adult tissues and expressed at high levels in kidney cortex and intestine. This indicates that the y<sup>+</sup>LAT1-4F2hc heterodimer, besides participating in amino acid uptake/secretion in many cell types, is the basolateral amino acid exchanger involved  
10 in transepithelial reabsorption of cationic amino acids; hence, its defect might be the cause of the human genetic disease lysinuric protein intolerance.

The gene encoding the disclosed cDNA is believed to reside on chromosome 14. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 14.

15 Preferred polypeptides comprise the following amino acid sequence:  
LALYSALFSYSGWDTLN (SEQ ID NO: 237), VTEEIKNPERNLPL (SEQ ID NO: 238), IGISMPIVT (SEQ ID NO: 239), IYILTNAVYYTVL (SEQ ID NO: 240), SDA  
VAVTFADQ (SEQ ID NO: 241), VALSCFGGLNASI (SEQ ID NO: 242), SRLFFV  
GSREGHLPD (SEQ ID NO: 243), SFSYWFFVGLS (SEQ ID NO: 244), VGQLYLR  
20 WKEP (SEQ ID NO: 245), RPRPLKLSVFFPIVFC (SEQ ID NO: 246), DTINSLIGI  
(SEQ ID NO: 247), LLAAACICLLTFINCA YVKWGT LVQDIFTYAKVLALIAVI  
VAGIVRLGQGASTHFENSFEGSSFAVGDI ALALYSALFSYSGWDTLNYVTEEI  
KNPERNLPLSIGISMPIVTIIYILTNAVYYTVLDMRDILASDAVAVTFADQIFGIF  
25 FNGIMALIYLCVEDIFQLINYYSFSYWFFVGLSIVGQLYLRWKEPDRPRPLKLS  
VFFPIVFCLCTIFLVAVPLYSDTINSLIGIAIALSGLPFYFLIIRVPEHKRPLYLRRI  
VGSATRYLQVLCMSVAAEMDLEDGGEMPKQRDPKSN (SEQ ID NO: 249)  
and/or ATALPPKIVGSATRYLQVLCMSVAAEMDLEDGGEMPKQRDPKSN  
(SEQ ID NO: 248). Polynucleotides encoding these polypeptides are also provided.

30 Contact of cells with supernatant expressing the product of this gene has been shown to increase the permeability of the plasma membrane of THP-1 monocyte cells to calcium. Thus, it is likely that the product of this gene is involved in a signal

transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of both THP-1 monocytes, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating monocytes.

5        This gene is expressed primarily in endothelial cells and brain, and, to a lesser extent, in a wide variety of human tissues.

         Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
10       not limited to, disorders of the neural or gastrointestinal systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulation system or central nervous system, expression of this gene at significantly higher or  
15       lower levels is routinely detected in certain tissues or cell types (e.g., neural, gastrointestinal, cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not  
20       having the disorder.

         Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 124 as residues: Glu-102 to Asn-110, Arg-256 to Leu-266, Pro-316 to Trp-328, Pro-331 to Arg-336, Met-350 to Gly-358.

Polynucleotides encoding said polypeptides are also provided.

25       The tissue distribution in brain combined with its homology to a amino acid transporter and biological activity of increasing ion flux in monocytes indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the  
30       "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease,



Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:11 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1550 of SEQ ID NO:11, b is an integer of 15 to 1564, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:11, and where b is greater than or equal to a + 14.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 2

The gene encoding the disclosed cDNA is believed to reside on the X chromosome. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for the X chromosome.

Preferred polypeptides of the invention comprise the following amino acid sequence:

AARGSGVRDPLEEAVCPFSDLQLHAGRTTALFKAVRQGHLSLQRLLLSFVCL  
CPAPRGGAYRGRQASLSCGGLHPVRASRLCLPKQAWAMAGAPPPVSLPPCS  
LISDCCASNQRDSVG (SEQ ID NO: 250). Polynucleotides encoding these  
polypeptides are also provided.

5        This gene is expressed primarily in cord blood cells, and, to a lesser extent, in  
frontal lobe of the brain.

         Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions which include, but are  
10    not limited to, developmental, reproductive, hematopoietic or neural disorders.  
Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
providing immunological probes for differential identification of the tissue(s) or cell  
type(s). For a number of disorders of the above tissues or cells, particularly of the  
immune and central nervous systems, expression of this gene at significantly higher or  
15    lower levels is routinely detected in certain tissues or cell types (e.g., immune,  
developmental, or cancerous and wounded tissues) or bodily fluids (e.g., lymph,  
amniotic fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue  
or cell sample taken from an individual having such a disorder, relative to the  
standard gene expression level, i.e., the expression level in healthy tissue or bodily  
20    fluid from an individual not having the disorder.

         Preferred polypeptides of the present invention comprise immunogenic  
epitopes shown in SEQ ID NO: 125 as residues: His-56 to Gln-65, Leu-80 to Ile-85.  
Polynucleotides encoding said polypeptides are also provided.

         The tissue distribution in cord blood cells indicates polynucleotides and  
25    polypeptides corresponding to this gene are useful for the treatment and diagnosis of  
hematopoietic related disorders such as anemia, pancytopenia, leukopenia,  
thrombocytopenia or leukemia since stromal cells are important in the production of  
cells of hematopoietic lineages. Representative uses are described in the "Immune  
Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19,  
30    20, and 27, and elsewhere herein. Briefly, the uses include bone marrow cell ex-vivo  
culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or  
chemotherapy of neoplasia.

The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Alternatively, polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus, this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:12 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1743 of SEQ ID NO:12, b is an integer of 15 to 1757, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:12, and where b is greater than or equal to a + 14.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene is expressed primarily in human T cell lymphomas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are not limited to, T cell lymphoma, immunodeficiencies, in addition to other immune system disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential  
5 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample  
10 taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 126 as residues: Met-1 to Phe-10. Polynucleotides  
15 encoding said polypeptides are also provided.

The tissue distribution in human T cell lymphomas indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14,  
20 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting  
25 immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory  
30 bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity

disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:13 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1359 of SEQ ID NO:13, b is an integer of 15 to 1373, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:13, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 4

The protein product of this clone shares sequence homology with the C-terminus of a human N-acetylglucosamine-phosphate mutase (See, e.g., Genbank Accession No. gb|AAC72409.1| (AF102265); all references available through this accession are hereby incorporated by reference herein.) Hofmann, et al. (Eur. J. Biochem. 221:741-747 (1994)) studied the N-acetylglucosamine-phosphate mutase of *Saccharomyces cerevisiae* and showed it to be essential for viability. A *S. cerevisiae* agm1 deletion mutant progressed through only approximately five cell cycles to form a 'string' of undivided cells with an abnormal cell morphology resembling

glucosamine auxotrophic mutants. Expression of the AGM1 gene on a multi-copy plasmid led to a significantly increased N-acetylglucosamine-phosphate mutase activity. Unlike over-expression of the *S. cerevisiae* AGM1 gene in a phosphoglucomutase (pgm1 delta/pgm2 delta) double deletion mutant which could  
 5 restore phosphoglucomutase activity, over-expression of the PGM2 gene encoding the major isoenzyme of phosphoglucomutase did not increase N-acetylglucosamine-phosphate-mutase activity and did not restore growth of agm1 deletion mutant cells. These observations indicate that the different hexosephosphate mutases of *S. cerevisiae* have partially overlapping substrate specificities but, nevertheless, distinct  
 10 physiological functions. The human N-acetylglucosamine-phosphate-mutase is expected to share at least some biological activities with the Agm1 protein.

Preferred polypeptide fragments of the invention comprise the following amino acid sequences: LSKAFLDSPNRLLAVEMNTDHLRLTVPNGIGALKLRXM EHYFSQGLSVQLFNDGSKGKLNHLGADFKSHQKPPQGMEIKSNERCCSFD  
 15 GDADRIVYYYHDADGHFHLIDGDKIATLISSFLKELLVEIGESLNIGVVQTAYA NGSSTRYLEEVMKVPVYCTKTGVKHLHHKAQEFDIGVYFEANGHGTALEST AVEMKIKQSAEQLEDKKRKAAMLENIIDLFNQAAGDAISDMLVIEAILALK GLTVQQWDALYTDLPNRQLKVQVADRRVISTTXAERQAVTPPGLQEAINDL VKKYKLSRAFVRPSGTEDVVRVYAEADSQESADHLAHEVSLAVFQLAGGIGE  
 20 RPQPGF (SEQ ID NO: 251), LSKAFLDSPNRLLAVEMNTDHLRLTV (SEQ ID NO: 252), PNGIGALKLRXMEHYFSQGLSVQLFNDG (SEQ ID NO: 253), SKGKL NHLGADFKSHQKPPQGMEIKS (SEQ ID NO: 254), NERCCSFDGDADRIV YYYHDADGHFHLI (SEQ ID NO: 255), DGDKIATLISSFLKELLVEIGESLNIGV (SEQ ID NO: 256), VQTAYANGSSTRYLEEVMKVPVYCTKTG (SEQ ID NO: 257), VKHLHHKAQEFDIGVYFEANGHGTALEST (SEQ ID NO: 258), TAVEMK  
 25 IKQSAEQLEDKKRKAAMLENI (SEQ ID NO: 259), IDLFNQAAGDAISDMLVIEAILALKGLT (SEQ ID NO: 260), VQQWDALYTDLPNRQLKVQVADRR VIST (SEQ ID NO: 261), TXAERQAVTPPGLQEAINDLVKKYKLSR (SEQ ID NO: 262), AFVRPSGTEDVVRVYAEADSQESA (SEQ ID NO: 263), and/or DH  
 30 LAHEVSLAVFQLAGGIGERPQPGF (SEQ ID NO: 264). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 6. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 6.

This gene is expressed primarily in fetal brain, and, to a lesser extent, in a wide variety of human tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental disorders, particularly of the central nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central and peripheral nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 127 as residues: Asn-36 to Lys-42, Lys-53 to Gln-60, Ile-64 to Ala-77, Ala-128 to Tyr-135, Lys-184 to Ala-199, Leu-245 to Leu-250. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution of N-acetylglucosamine-phosphate mutase in fetal brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital

malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition,  
5 elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity,  
10 to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly  
15 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:14 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
20 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3726 of SEQ ID NO:14, b is an integer of 15 to 3740, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:14, and where b is greater than or equal to a + 14.

## 25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 5**

This gene is expressed primarily in human stomach and stomach tumor cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
30 not limited to, disorders of the gastrointestinal system, particularly cancer or ulcers of stomach tissue. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the



tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., gastrointestinal, or cancerous and wounded tissues) or bodily fluids (e.g., bile, lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in tumors of the stomach indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, treatment and intervention of these tumors, in addition to other tumors where expression has been indicated. Additionally, the protein product of this gene may play a role in the normal function of the stomach and/or digestive system. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:15 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1182 of SEQ ID NO:15, b is an integer of 15 to 1196, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:15, and where b is greater than or equal to a + 14.

### **30 FEATURES OF PROTEIN ENCODED BY GENE NO: 6**

Preferred polypeptides of the invention comprise the following amino acid sequences:

FEIALPRESNITVLIKLGTPDLLAKPCYIVISKRHITMLSIKSGERIVFTFSCQSPE  
NHFVIEIQKNIDCMMSGPCPFGEVQLQPSTSLPTLNRTFIWDVKAHKSIGLELQ  
FSIPRLRQIGPGESCPDGVTHSISGRIDATVVRIGTFCSNGTVSRIKM (SEQ ID  
NO: 266), and/or GTRAAPGLGAWGRRSPPSFSPRRPGVMAGLNCVSIAL  
5 LGVLLLGAARLPRGAFAFEIALPRESNITVLIKLGTPDLLAKPCYIVISKRHITM  
LSIKSGERIVFTFSCQSPENHFVIEIQKNIDCMMSGPCPFGEVQLQPSTSLPTLNRT  
FIWDVKAHKSIGLELQFSIPRLRQIGPGESCPDGVTHSISGRIDATVVRIGTFC  
SNGTVSRIKMQEGVKMALHLPWFHPRNVSGFSIANRSSIKRLCIIESVFEGEGS  
ATLMSANYPEGFPEDELMTWQFVVPAPHLRASVSFLNFNLSNCERKEERVEYY  
10 IPGSTTNPEVFKLEDKQPGNMAGNFNLSLQGCDQDAQSPGILRLQFQVLVQH  
PQNESNKIYVVDLSNERAMSLTIEPRPVKQSRKFVPGCFVCLESRTCSSNLTLT  
SGSKHKISFLCDDLTRLWMNVEKP (SEQ ID NO: 265). Polynucleotides encoding  
these polypeptides are also provided.

15 This gene is expressed primarily in placenta, and to a lesser extent in, prostate  
and ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions which include, but are  
not limited to, male and female infertility, and associated disorders of the  
20 reproductive system. Similarly, polypeptides and antibodies directed to these  
polypeptides are useful in providing immunological probes for differential  
identification of the tissue(s) or cell type(s). For a number of disorders of the above  
tissues or cells, particularly of the reproductive system, expression of this gene at  
significantly higher or lower levels is routinely detected in certain tissues or cell types  
25 (e.g., reproductive, or cancerous and wounded tissues) or bodily fluids (e.g., lymph,  
amniotic fluid, seminal fluid, serum, plasma, urine, synovial fluid and spinal fluid) or  
another tissue or cell sample taken from an individual having such a disorder, relative  
to the standard gene expression level, i.e., the expression level in healthy tissue or  
bodily fluid from an individual not having the disorder.

30 The tissue distribution of this gene in the prostate, placenta and ovary  
indicates that polynucleotides and polypeptides corresponding to this gene are useful  
for treatment, prevention, and/or diagnosis of male or female infertility, endocrine

disorders, fetal deficiencies, ovarian failure, amenorrhea, ovarian cancer, benign prostate hyperplasia and prostate cancer. Similarly, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression within placental tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus, this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:16 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2195 of SEQ ID NO:16, b is an integer of 15 to 2209, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:16, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 7

The translation product of this gene shares homology with the human and rat HNK-1 sulfotransferase protein (See, e.g., Genbank Accession Nos. gb|AAB88123.1| (AF022729) and gil2921306|gb|AAC04707.1| (AF033827); all references available through these accessions are hereby incorporated herein by reference.) Ong E, et al. (J Biol Chem. 273(9):5190-5 (1998)) have characterized the human HNK-1 sulfotransferase, and show that it is involved in the synthesis of the HNK-1

carbohydrate epitope which is expressed on various adhesion molecules in the nervous system and on immune cells (e.g., natural killer cells) and is suggested to play a role in cell-cell and cell-substratum interactions. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with HNK-1 sulfotransferase proteins. Such activities are known in the art, some of which are described elsewhere herein, or in, for example, Bakker, et al., J Biol Chem. 272:29942-6 (1997), incorporated herein by reference. Based on sequence similarity between sulfotransferases, a consensus sequence for the active site was developed (Ong, et al., supra). The consensus pattern is as follows:

xxRPDzzzz, where x represents hydrophobic amino acid residues and z represents any amino acid residue. Therefore,

Preferred polypeptides of the invention comprise the following amino acid sequences: FVRDPFVRL (SEQ ID NO: 267), FLFVRDPFVRLIS (SEQ ID NO: 268), FLFVRDPFVRLISAF (SEQ ID NO: 269), and/or YLHTSFSRPHTGPPLPTPG PDRDRELTADSDVDEFLDKFLSAGVKQSDLPRKETEQPPAPGSMEENVRGY DWSPRDARRSPDQGRQQAERRSVLRGFCANSSLAFTKERAFDDIPNSELSHL IVDDRHGAIYCYVPKVACTNWKRVMI VLSGSL LHRGAPYRDPLRIPREHVH NASAHLTFNKFWRRYGKLSRHLMKVKLKKYTKFLFVRDPFVRLISAFRSK FELENEEFYRKFAVPMLRLYANHTSLPASAREAFRAGLKVSFANFIQYLLDPH TEKLAPFNEHWRQVYRLCHPCQIDYDFVGKLETLDDEAAQLLQLLQVDRQ LRFPPSYRNRTASSWEEDWFAKIPLAWRQQLYKLYEADFVLFGYPKPENLL RD (SEQ ID NO: 270). Polynucleotides encoding these polypeptides are also provided.

Further preferred are the sulfotransferase active site polypeptides listed above, and at least 5, 10, 15, 20, 25, 30, 50, or 75 additional contiguous amino acid residues of the sequence referenced in Table I for this gene. The additional contiguous amino acid residues is N-terminal or C-terminal to the sulfotransferase active site polypeptides. Alternatively, the additional contiguous amino acid residues is both N-terminal and C-terminal to the sulfotransferase active site polypeptides, wherein the total N- and C-terminal contiguous amino acid residues equal the specified number. The above preferred polypeptide domains are characteristic of a signature specific to sulfotransferase proteins. The nucleotides sequence of this gene was found to be

homologous to the human hypoxanthine guanine phosphoribosyl transferase 2 cDNA which is known to be involved in the purine salvage pathway resulting in the maintenance of homeostatic levels of uric acid (See Genbank Accession No.T30127).

The gene encoding the disclosed cDNA is believed to reside on chromosome 7. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 7.

This gene is expressed to a very high level in HL-60 myelogenous leukemia cell lines, and to a lesser extent, in most cell types of the immune system.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune, myelopoiesis, and metabolic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, metabolic, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 130 as residues: Ser-39 to Gly-46, Leu-49 to Ala-62, Lys-79 to Ala-93, Gly-95 to Asp-105, Ser-107 to Val-127, Gly-193 to Leu-200, Lys-218 to Ser-227, Lys-234 to Thr-239, Pro-366 to Asp-379, Pro-406 to Asp-414. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in HL-60 myelogenous leukemia cell lines and homology to HNK-1 sulfotransferase proteins indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, prevention and/or treatment of a variety of immune system disorders, including but not limited to, those involving the HNK-1 carbohydrate epitope, (e.g. HNK-1 as an auto-antigen

in peripheral demyelinating neuropathy). Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, expression of this gene product in tonsils indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Alternatively, the homology to a conserved purine metabolism protein may suggest that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, prevention, and/or treatment of various metabolic disorders such as Tay-Sach's Disease, phenylketonuria, galactosemia, porphyrias, Hurler's syndrome, and various urogenital disorders related to metabolic conditions, particularly Lesch-Nyhan syndrome. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:17 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically  
5 excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1760 of SEQ ID NO:17, b is an integer of 15 to 1774, where both a and b correspond to the positions of nucleotide  
10 residues shown in SEQ ID NO:17, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 8

When tested against Jurkat T-cell lines, supernatants removed from cells containing this gene activated the gamma activating sequence (GAS) promoter  
15 element. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway, a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. Thus, it is likely that  
20 this gene activates T-cells through the Jak-STAT signal transduction pathway.

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: KLVRLQVPVRNSRVDPRVRSKIGSRRWMLQLI  
25 MQLGSVLLTRCPFWGCFSQLMLYAERAEARRKPDIPVPYLYFDMGAAVLCA  
SFMSFGVKRRWFALGAALQLAISTYAA YIGGYVHYGDWLKVRMYSRTVAII  
GGFLVLASGAGELYRRKPRSRLQSTGQVFLGIYLICVAYSLQHSKEDRLA  
YLNHLPGGELMIQLFFVLYGILALAFSLGYVYVTLAAQILAVLLPPVMLLIDG  
NVAYWHNTRRVEFWNQMKLLGESVGIFGTAVILATDG (SEQ ID NO: 271).

30 A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MQLGSVLLTRCPFWGCFSQLMLYAERAEARRKPDIPVP  
YLYFDMGAAVLCASFMSFGVKRRWFALGAALQLAISTYAA YIGGYVHYGD

WLKVRMYSRTVAIIGGFLVLASGAGELYRRKPRSRSLQSTGQVFLGIYLICVA  
YSLQHSKEDRLAYLNHLPGGELMIQLFFVLYGILAPGLSVRLLRDPRCPDPGC  
TAAPCHAAH (SEQ ID NO: 272). Polynucleotides encoding these polypeptides are also provided.

5           The gene encoding the disclosed cDNA is believed to reside on chromosome 17. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 17.

          This gene is expressed primarily in endometrial tumors, and to a lesser extent, in T-cells, pituitary and to a certain extent in most cell types.

10           Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, female reproductive, immune, or endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing  
15 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and/or immune systems expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, reproductive, or cancerous and wounded tissues) or bodily fluids (e.g., amniotic fluid, lymph, serum,  
20 plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

          Preferred polypeptides of the present invention comprise immunogenic  
25 epitopes shown in SEQ ID NO: 131 as residues: Ala-27 to Asp-34, Tyr-116 to Leu-125. Polynucleotides encoding said polypeptides are also provided.

          The tissue distribution predominantly in the endometrium indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of a range of immune and/or reproductive  
30 disorders including endometriosis, endometritis, and endometrioma. Similarly, the tissue distribution in T-cells and the ability of supernatants expressing this gene to stimulate the GAS promoter element in T-cells indicates polynucleotides and



polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted

factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Alternatively, the tissue distribution in pituitary indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", and "Binding Activity" sections below, in Example 11, 17, 18, 19, 20 and 27, and elsewhere herein. Briefly, the protein can be used for the detection, treatment, and/or prevention of the Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g., diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-

hypopituitarism), thyroid (e.g., hyper-, hypothyroidism), parathyroid (e.g., hyper-, hypoparathyroidism), hypothalamus, and testes.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are  
5 related to SEQ ID NO:18 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
10 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1660 of SEQ ID NO:18, b is an integer of 15 to 1674, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:18, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 9

15 Contact of cells with supernatant expressing the product of this gene has been shown to increase the permeability of the plasma membrane of the myeloid leukemia cell line AML-193 to calcium. Thus, it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a  
20 receptor on the surface of the plasma membrane of myeloid leukemia cells, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating myeloid leukemia cells.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the  
25 following amino acid sequence: SNEILLSFPQNYIQLNGSLIHGLWNLASLFS  
NLCLFVLMPPAFFFLESEGFAGLKKGIRARILETLVMLLLLALLILGIVWVAS  
ALIDNDAASMESLYDLWEFYLPYLYSCISLMGCLLLLCTPVGLSRMFTVMG  
HLLVKPTILEDLDEQIYIITLEEEALQRRNLGLSSSVVEYNIMELEQELENVKTL  
KTKLERRKKASAWERNLVYPAVMVLLLIETSSISVLLVACNILCLLVDETAM  
30 PKGTRGPGIGNASLSTFGFVGAALHILIFYLMVSSVVGIFYSLRFFGNFTPKKD  
DTMTKIIGNCVSILVLSSALPVMSRTLGITRFDLLGDFGRFNWLGNFYIVLS  
YNLLFAIVTTLCLVRKFTSAVREELFKALGLHKLHLPNTSRDSETAKPSVNGH

QKAL (SEQ ID NO: 273). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in fetal heart, and to a lesser extent, in colon and the adult pulmonary system.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, heart, lung and digestive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological  
10 probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular, pulmonary and digestive systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., developmental, cardiovascular, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid,  
15 serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic  
20 epitopes shown in SEQ ID NO: 132 as residues: Glu-67 to Asn-74, Glu-88 to Asn-93, Lys-95 to Ala-107, Ala-147 to Arg-153, Phe-197 to Thr-205, Pro-292 to His-308. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in fetal heart, colon and pulmonary tissues and the biological activity in increasing the permeability of the plasma membrane of the  
25 myeloid leukemia cell line AML-193 to calcium, likely indicating that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of myeloid leukemia cells, indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment, prevention, and/or detection of a range of disorders including  
30 a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, myocardial infarction, myocarditis, ischemia, thrombosis, coronary artery disease,

arteriosclerosis, and/or atherosclerosis; pulmonary edema and embolism, bronchitis and/or cystic fibrosis; Crohn's Disease and/or colon cancer. Similarly, the tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:19 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2004 of SEQ ID NO:19, b is an integer of 15 to 2018, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:19, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 10

The protein product of this clone shares sequence homology with the human MaxiK channel beta 2 subunit (See, Genbank Accession No. gblAAD23380.11AF099137\_1 (AF099137); all references available through this accession are hereby incorporated herein by reference), which is believed to be a modulatory subunit of the voltage and Ca<sup>2+</sup> activated K<sup>+</sup> (MaxiK) channel. Additionally, this protein shares homology to the human calcium-activated potassium

channel beta subunit, which, when combined with its corresponding alpha subunit and modulating peptide, are believed to be useful in treating asthma, angina, hypertension, incontinence, migraine, irritable bowel syndrome (IBS). The subsequent heteromultimer that forms upon combining the alpha, beta, and modulator subunits  
5 are also thought to be useful in preventing premature labour, preventing and treating cerebral ischemia, inducing pain modulation and decreasing neurogenic inflammation in a patient (See GeneSeq Accession No. R85306).

Preferred polypeptides comprise the soluble domain which consists of the following amino acid sequence: RSYMQSVWTEESQCTLLNASITETFNCSFSCGP  
10 DCWKLSQYPCLQVYVNLTSSEKLLLYHTEETIKINQKCSYIPKCGKNFEESM  
SLVNVVMENFRKYQHFSCYSDEGNQKSVILTKLYSSNVLFHSLFWPTCMMAGGVAIVAMVKLTQYLSLLCERIQRINR (SEQ ID NO: 274). Polynucleotides encoding these polypeptides are also provided. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities  
15 with modulatory subunits of voltage and Ca<sup>2+</sup> activated K<sup>+</sup> channel proteins. Such activities are known in the art, some of which are described in Wallner, et al., PNAS 96:4137-4142 (1999), incorporated herein by reference.

This gene is expressed primarily in adrenal gland tumor, and to a lesser extent, in Hodgkin's lymphoma.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, endocrine and immune disorders, particularly Hodgkin's Lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
25 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and/or endocrine systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, endocrine, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum,  
30 plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression

level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 133 as residues: Trp-25 to Gln-30, Pro-50 to Gln-57,  
5 Pro-93 to Glu-101, Arg-114 to Cys-121, Ser-123 to Gln-129, Ile-177 to Arg-182. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in adrenal gland tumor and its identification as the modulatory subunit of the voltage and  $\text{Ca}^{2+}$  activated  $\text{K}^{+}$  (MaxiK) channel indicates that polynucleotides and polypeptides corresponding to this gene are useful for the  
10 detection, treatment, and/or prevention of various endocrine disorders and cancers, particularly Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g., diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g., hyper-, hypothyroidism), parathyroid (e.g., hyper-, hypoparathyroidism), hypothalamus, and testes. Alternatively, expression in  
15 proliferative immune tissues combined with its homology to a novel human K channel indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein.  
20 Briefly, the expression of this gene product in Hodgkin's lymphoma indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune  
25 responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory  
30 bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity

disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:20 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2084 of SEQ ID NO:20, b is an integer of 15 to 2098, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:20, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 11

The translation product of this gene shares homology with collagen and collagen like proteins (See, e.g., Genbank Accession Nos. gil2920535|gb|AAC39658.11 (AF018081) and gil2384942|gb|AAB69961.11 (AF022985); all references available through these accession numbers are hereby incorporated by reference herein). Additionally, it has been determined that this gene has homology to the human Kruppel related zinc finger protein (HTF10) which is known to be important as a transcription factor, particularly in development (See Genebank Accession No.L11672).

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: AFAHLQLGPMWKLWRAEEGAAALGGALFLLL  
5 FALGVRQLLKQRRPMGFPPGPPGLPFIGNIYSLAASSELPHVYMRKQSQVYG  
EVQPRRAPGREGRQAGPGWPGPSWLDLWPPLGRLVGTSPCAGCPLRDTRFPG  
LEGRS PRRRAPLQGEPRPCR (SEQ ID NO: 275). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in human erythroleukemia cell line (HEL),  
10 serum induced smooth muscle, and to a lesser extent in human 8 week whole embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, leukemia, musculoskeletal, or developmental disorders. Similarly,  
15 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic system and muscular system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune,  
20 musculoskeletal, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 134 as residues: Leu-30 to Gly-38, Arg-67 to Val-72, Val-76 to Ala-89, Pro-118 to Arg-123, Gly-129 to Ala-136, Leu-138 to Arg-146. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in human erythroleukemia cell line (HEL), and serum  
30 induced smooth muscle, and the shared homology with collagen and collagen like proteins indicates that polynucleotides and polypeptides corresponding to this gene are useful for disorders of hematopoietic or muscular systems, such as leukemia and



muscular dystrophy. Additionally, the shared homology with collagen proteins would suggest that this protein may also be important in the diagnosis or treatment of various autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, dermatomyositis, etc.); dwarfism, spinal deformation, joint abnormalities, and  
5 chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid, etc.).

The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents  
10 that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines;  
immunostimulating/immunosuppressant activities (e.g., for treating human  
15 immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g., for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g., for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g., for treating  
20 infections, tumours); haemostatic or thrombolytic activity (e.g., for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g., for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy  
25 procedures. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

30 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:21 and may have been publicly available prior to conception of

the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general  
5 formula of a-b, where a is any integer between 1 to 1732 of SEQ ID NO:21, b is an integer of 15 to 1746, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:21, and where b is greater than or equal to a + 14.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 12

10 A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MRVRIGLTLLCAVLLSLASASSDEEGSQD  
ESLGFQDYFDIR (SEQ ID NO: 276). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome  
15 8. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 8.

This gene is expressed primarily in dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at  
25 significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
30 individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 135 as residues: Ser-22 to Ser-41, Glu-43 to Thr-50,

Ser-63 to Leu-68, Ser-71 to Gly-84, Ser-96 to Gly-114. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in dendritic cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis, prevention, and/or treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types.

The secreted protein can be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation

modulating activity or induction of other cytokines:

immunostimulating/immunosuppressant activities (e.g., for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy);

regulation of haematopoiesis (e.g., for treating anaemia or as adjunct to

5 chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g., for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g., for treating infections, tumours); haemostatic or thrombolytic activity (e.g., for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g., for treating  
10 septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease: for regulation of metabolism, behaviour, and many others.

Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to  
15 identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are  
20 related to SEQ ID NO:22 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general  
25 formula of a-b, where a is any integer between 1 to 2862 of SEQ ID NO:22, b is an integer of 15 to 2876, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:22, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 13

30 A preferred polypeptide variant of the invention comprises the following amino acid sequence: MARGSLRRLRLVLGLWLALLRSVAGEQAPGTAPC SRGSSWSADLDKCMDCSTSCPLPA ALAHPWGRSEPDRLAGAAFWLFGLE

TMPQE REVHHPHRGDRRRGLPSCGADPVTMCPLPAGARPLIIHSSILEPVSAS  
QTRREPSSSNHK GGGGR (SEQ ID NO: 277). Polynucleotides encoding these  
polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome  
5 16. Accordingly, polynucleotides related to this invention are useful as a marker in  
linkage analysis for chromosome 16.

This gene is expressed primarily in tumor growth factor or lipopolysaccharide  
treated bone marrow stroma, epithelioid sarcoma, umbilical vein endothelial cells, and  
to a lesser extent, in other tissues.

10 Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions which include, but are  
not limited to, hematopoiesis or immune disorders. Similarly, polypeptides and  
antibodies directed to these polypeptides are useful in providing immunological  
15 probes for differential identification of the tissue(s) or cell type(s). For a number of  
disorders of the above tissues or cells, particularly of the hematopoietic,  
integumentary, or immune systems expression of this gene at significantly higher or  
lower levels is routinely detected in certain tissues or cell types (e.g., hematopoietic,  
immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum,  
20 plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken  
from an individual having such a disorder, relative to the standard gene expression  
level, i.e., the expression level in healthy tissue or bodily fluid from an individual not  
having the disorder.

Preferred polypeptides of the present invention comprise immunogenic  
25 epitopes shown in SEQ ID NO: 136 as residues: Pro-35 to Trp-42, Pro-65 to Asp-72,  
Thr-86 to Phe-93, Ile-97 to Glu-103. Polynucleotides encoding said polypeptides are  
also provided.

The tissue distribution in tumor growth factor or lipopolysaccharide treated  
bone marrow stroma, epithelioid sarcoma, and umbilical vein endothelial cells  
30 indicates that polynucleotides and polypeptides corresponding to this gene are useful  
for the diagnosis, prevention, and/or treatment of a variety of immune system  
disorders. Representative uses are described in the "Immune Activity" and "infectious

disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types.

The secreted protein can be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g., for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g., for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g., for treating wounds, stimulation of follicle stimulating hormone (for

control of fertility); chemotactic and chemokinetic activities (e.g., for treating infections, tumours); haemostatic or thrombolytic activity (e.g., for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g., for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other  
5 hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly  
10 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:23 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
15 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1038 of SEQ ID NO:23, b is an integer of 15 to 1052, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:23, and where b is greater than or equal to a + 14.

---

## 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 14

The translation product of this gene shares sequence homology with chromaffin granule amine transporter protein which is thought to be important in vesicle membrane amine transport, particularly in the neural and endocrine tissue, and the human vesicular monoamine transporter hVMAT1 which is involved in the  
25 regulation of amine storage in cardiovascular, endocrine, and central nervous system function (See, Genbank Accession Nos. gil1314290 and gblAAC50472.1; all references available through these accession numbers are hereby incorporated by reference herein). Based on these sequence similarities, The translation product of this gene is expected to share at least some biological activities with amine transporter  
30 proteins. Such activities are known in the art, some of which are described in Erickson, et al., PNAS 93:5166-5171 (1996), and/or Liu, et al., Cell 70:539-551 (1992), which are both incorporated herein by reference.

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: GTSFLDPTLSLFVLEKFNLPAGYVGLVFLGMAL  
5 SYAISSPLFGLLSDKRPPLRKWLLVFGNLITAGCYMLLGPVPILHIKSQWLWLL  
VLILVVSGLSAGMSIIPTFPEILSCAHENGFEGLSTLGLVSGLFSAMWSIGAF  
MGPTLGGFLYEKIGFEWAAAIQGLWALISGLAMGLFYLLLEYSRRKRKRSKSNIL  
STEEERTTLLPNET (SEQ ID NO: 278). Polynucleotides encoding these  
polypeptides are also provided.

10 The gene encoding the disclosed cDNA is believed to reside on chromosome 6. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 6.

This gene is expressed primarily in colon cancer, osteoclastoma, and T-cell lymphoma, and to a lesser extent in many tumor or proliferative tissues such as  
15 endometrial tumor, chondrosarcoma, induced umbilical vein endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diseases resulting from disorders in small molecule transport (i.e.,  
20 signalling molecules) in afflicted tissues and organs, particularly of the endocrine and central nervous systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the musculoskeletal, immune, and/or digestive systems  
25 and cancer expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, endocrine, or cancerous and wounded tissues) or bodily fluids (e.g., bile, lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression  
30 level in healthy tissue or bodily fluid from an individual not having the disorder.



Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 137 as residues: Ser-114 to Asn-123, Thr-127 to Thr-132. Polynucleotides encoding said polypeptides are also provided.

5 The tissue distribution in colon cancer, osteoclastoma, and T-cell lymphoma and homology to amine transporter family members indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders or diseases resulted from small molecule transport in afflicted tissues and organs, particularly that of colon, osteoclast or T-cells. The expression in cancer tissues, and shared homology with transporter proteins may also indicate its role in  
10 anti-cancer drug resistance. Additionally, the protein can be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell  
15 proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g. for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or  
20 nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumours); haemostatic or thrombolytic activity (e.g. for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other  
25 hyperproliferative disease; or for identifying inhibitors or promoters of the transport of toxic molecules to vesicles, for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

30 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:24 and may have been publicly available prior to conception of

the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general  
5 formula of a-b, where a is any integer between 1 to 1527 of SEQ ID NO:24, b is an integer of 15 to 1541, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:24, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 15

10 The translation product of this gene shares sequence homology with the human prolyl 4-hydroxylase alpha (II) subunit which is important in catalyzing the formation of 4-hydroxyproline in collagens which is essential for the folding of newly synthesised collagen polypeptide chains into triple-helical molecules (See Genbank Accession No. gblAAB71339.11; all references available through this accession are  
15 hereby incorporated herein by reference). Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with Prolyl 4-hydroxylase proteins. Such activities are known in the art, some of which are described in Annunen, et al., J. Biol. Chem. 272:17342-17348 (1997) which is incorporated herein by reference.

20 When tested against U937 myeloid and Jurkat T-cell cell lines, supernatants removed from cells containing this gene activated the gamma activating sequence (GAS), a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the  
25 Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. Thus, it is likely that this gene activates myeloid cells and T-cells through the Jak-STAT signal transduction pathway.

In a specific embodiment, polypeptides comprising the amino acid sequence  
30 of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence:

GTREARLRDLTRFYDKVLSLHEDSTTPVANPLLAFTLIKRLQSDWRNVVHSL  
EASENIRALKDGYEKVEQDLPAFEDLEGAARALMRLQDVYMLNVKGLAR  
GVFQRVGTGSAITDLYSPKRLFSLTGDDCFQVGKVAYDMGDYYHAIPWLEEA  
VSLFRGSYGEWKTEDEASLEDALDHLAFAYFRAGNVSCALSLSREFLLYSPD  
5 NKRMAARNVLKYERLLAESPNHVVAEAVIQRPNIPHLQTRDITYEGLCQTL  
GSQPTLYQIPSLYCSYETNSNAYLLLQPIRKEVIHLEPYIALYHDFVSDSEAQ  
KIRELAEPWLQRSVVASGEKQLQVEYRISKSAWLKDTVDLKLVTLNHRIA  
LTGLDVRPPYAEYLQVVNYGIGGHYEPHFDHATSPSSPLYRMKSGNRVATFM  
IYLSSVEAGGATAFIYANLSVPVVRNAALFWWNLHRSGEGDSDTLHAGCP  
10 VLVGDKWVANKWIEHYGQEFRRPCSSSPED (SEQ ID NO: 282). Additional,

Preferred polypeptides comprise the following amino acid sequence: GTREA  
RLRDLTRFYDKVLSLHEDSTTPVANPLLAFTLIKRLQSDWRNVVHSLEASENI  
RALKDGYEKVEQ DLPAFEDLEGAARAL (SEQ ID NO: 279), ALMRLQD (SEQ  
ID NO: 280), and/or VEAGGAT (SEQ ID NO: 281). Polynucleotides

15 encoding these polypeptides are also provided.

This gene is expressed primarily in lymph node breast cancer, colon  
carcinoma, and to a lesser extent in osteoblasts and adipocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a

20 biological sample and for diagnosis of diseases and conditions which include, but are  
not limited to, disorders of connective and immune tissues, particularly autoimmune  
disorders. Similarly, polypeptides and antibodies directed to these polypeptides are  
useful in providing immunological probes for differential identification of the  
tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,  
25 particularly of the connective tissues in breast, colon, bone, and fat, expression of this  
gene at significantly higher or lower levels is routinely detected in certain tissues or  
cell types (e.g., immune, connective, or cancerous and wounded tissues) or bodily  
fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another  
tissue or cell sample taken from an individual having such a disorder, relative to the  
30 standard gene expression level, i.e., the expression level in healthy tissue or bodily  
fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 138 as residues: Ser-74 to Ala-84, Gln-156 to Tyr-161, Tyr-184 to Asn-189, Ser-218 to Ile-223, Pro-299 to Ser-308, His-359 to Thr-368, Tyr-390 to Asp-404. Polynucleotides encoding said polypeptides are also provided.

5       The tissue distribution in lymph node breast cancer and colon carcinoma and homology to prolyl 4-hydroxylase alpha (II) subunit indicates that polynucleotides and polypeptides corresponding to this gene are useful for intervention of connective tissue disorders and diseases (e.g. arthritis, trauma, tendonitis, chondromalacia and inflammation), as well as, in the diagnosis or treatment of various autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis as well as dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias ie. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid. Alternatively, the tissue distribution within various tissue carcinomas and tumor tissues, and biological activity reflected by the binding and activation of the GAS promoter element indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders.

15       Expression in cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

25       Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:25 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2065 of SEQ ID NO:25, b is an

30

integer of 15 to 2079, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:25, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 16

5 In an additional embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: IQPSHAALLHCRSTFRKTECLDPW  
10 WVRRQLLGMAGIGGLQKMKAPHTGVLHLGSVWVFLGPFLGVG YTLTFNPL  
SGCMSTVRWLNSNITANRTLRSVCHVTPLHRSLSPHDGEYLRQMLLNSSSR  
AGEAGSWG Y (SEQ ID NO: 283). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 20. Accordingly, polynucleotides related to this invention are useful as a marker in  
15 linkage analysis for chromosome 20.

This gene is expressed primarily in fetal liver, and, to a lesser extent, in a variety of fetal and other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, liver disorders and cancers (e.g., hepatoblastoma, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential  
25 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the liver, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., hepatic, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, bile, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an  
30 individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 139 as residues: Ser-67 to Tyr-75. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in fetal liver indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of liver disorders and cancers (e.g., hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:26 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1933 of SEQ ID NO:26, b is an integer of 15 to 1947, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:26, and where b is greater than or equal to a + 14.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 17**

The translation product of this gene shares sequence homology with human laminin B1 which is thought to be an important structural extracellular matrix component involved in cell migration and signalling, particularly in stimulating epithelial cell growth and differentiation (See, Genbank Accession No g1186837).

Preferred polypeptides of the invention comprise the following amino acid sequences: CSSPPGRLPWCWTAPRTLKGHGLISTLRLTAPLHLAWKMMLS RKALFVLLNTPVLFHALEGRFLFSKLCHHHTIQRTLTVPKFRSS (SEQ ID NO: 284), RSPTS RVQLLKRQSCPCQRNDLNEEPQHFTHYAIYDFIVKGSCFCNG  
 5 HADQCIPVHGFRPVKAPGTFHMHGKCM (SEQ ID NO: 285), and/or HNTAG SHCQHCAPLYNDRPWEAADGKTGAPNECRTCKCNGHADTCHFDVNVWEAS GNRSGGVCDDCQHNTGQYCQRCKPGFYRDLRRPFSAPDACKPCSCHPV GSAVLPANSVTFCDPNSGDCPCKPGVAGRRCDRCMVGYWGF GDYGCRP CDCAGSCDPITGDCISSHTDIDWYHEVPDFRPVHNKSEPAWEWEDAQGFSAL  
 10 LHSGKCECKEQTLGNAKAFCGMKYSYVLKIKILSAHDKGTHVEVNVKIK KVLKSTKLKIFRGKANIISRIMDGQ RMHLSNPQSWFGIPCSRT (SEQ ID NO: 286). Polynucleotides encoding these polypeptides are also provided.

Included in this invention as preferred domains are Laminin-type EGF-like (LE) domain signatures, which were identified using the ProSite analysis tool (Swiss  
 15 Institute of Bioinformatics). Laminins are the major noncollagenous components of basement membranes that mediate cell adhesion, growth migration, and differentiation. They are composed of distinct but related alpha, beta and gamma chains. The three chains form a cross-shaped molecule that consist of a long arm and three short globular arms. The long arm consists of a coiled coil structure contributed  
 20 by all three chains and cross-linked by interchain disulfide bonds. Beside different types of globular domains each subunit contains, in its first half, consecutive repeats of about 60 amino acids in length that include eight conserved cysteines. The tertiary structure of this domain is remotely similar in its N-terminal to that of the EGF-like module. It is known as a 'LE' or 'laminin-type EGF-like' domain. The number of  
 25 copies of the LE domain in the different forms of laminins is highly variable; from 3 up to 22 copies have been found. A schematic representation of the topology of the four disulfide bonds in the LE domain is shown below.

10

15

20

30

Further preferred are polypeptides comprising the laminin-type EGF-like domains listed above, and at least 5, 10, 15, 20, 25, 30, 50, or 75 additional contiguous amino acid residues of the sequence encoded by this gene. The additional contiguous amino acid residues is N-terminal or C-terminal to the laminin-type EGF-



like domain. Alternatively, the additional contiguous amino acid residues is both N-terminal and C-terminal to the laminin-type EGF-like domain, wherein the total N- and C-terminal contiguous amino acid residues equal the specified number. The above preferred polypeptide domain is characteristic of a signature specific to

5 Laminin proteins. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with Laminin proteins. Such activities are known in the art, some of which are described elsewhere herein.

This gene is expressed primarily in osteoblastic tissues and cell types, including osteoblasts, osteoblastomas and osteoclastomas. Expression is also

10 abundant in vascular-pulmonary tissues such as lung, micro-vasculature, pulmonary, endothelial and smooth muscle cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are

15 not limited to, cancer and malignancies (particularly of osteoblastic tissues and rhabdomyosarcoma), as well as cardiovascular and respiratory or pulmonary disorders such as athsma, pulmonary edema, pneumonia, atherosclerosis, restenosis, stoke, thrombosis hypertension, inflammation and wound healing. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological

20 probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardio-respiratory system, and skeletal system expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., skeletal, osteoblast, cardio-respiratory, vascular, or cancerous and wounded tissues) or bodily fluids (e.g., lymph,

25 pulmonary surfactant, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic

30 epitopes shown in SEQ ID NO: 140 as residues: Ser-28 to Cys-34, Thr-51 to Thr-58, Tyr-64 to Asn-81, Asp-111 to Lys-116, Asp-145 to Phe-160, Pro-203 to Glu-217. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in osteoblastic tissues and cell types and homology to laminin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment, prevention and diagnosis of cardiovascular and respiratory or pulmonary disorders such as asthma, pulmonary edema, pneumonia, atherosclerosis, restenosis, stroke, angina, thrombosis hypertension, inflammation and wound healing. As a homolog of laminin, this gene product quite possibly has a role in cell adhesion, migration, proliferation, angiogenesis, chondrogenesis, wound healing and oncogenesis. An EST (Int J Cancer 1996 May 16;66(4):571-577) with an identical sequence to part of this contig was shown to be differentially expressed in human primary myoblasts and embryonal rhabdomyosarcoma and therefore might have an important role in the determination or maintenance of the normal phenotype, and thus its loss is possibly involved in the progression of malignancies, particularly of skeletal muscle. Similarly, the homology to a laminin would suggest a role in the detection and treatment of disorders and conditions afflicting connective tissues (e.g. arthritis, trauma, tendonitis, chondromalacia and inflammation) in the diagnosis or treatment of various autoimmune disorders such as rheumatoid arthritis, lupus, scleroderma, and dermatomyositis as well as dwarfism, spinal deformation, and specific joint abnormalities as well as chondrodysplasias i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:27 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

formula of a-b, where a is any integer between 1 to 3365 of SEQ ID NO:27, b is an integer of 15 to 3379, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:27, and where b is greater than or equal to a + 14.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 18

The gene encoding the disclosed cDNA is believed to reside on chromosome 10. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 10.

10 In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: NISSQYCILKSLEMMISGLKLLVLFLKFAPENY CLSTETLQMPNRHLRLSKATCYLMKCLLP SYFE (SEQ ID NO: 289).

Polynucleotides encoding these polypeptides are also provided.

15 This gene is expressed primarily in placenta, brain, and to a lesser extent, in a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive, behavioral, or nervous system disorders, such as: depression, schizophrenia, Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, dementia, paranoia, addictive behavior, epilepsy, transmissible spongiform encephalopathy (TSE), Creutzfeldt-Jakob disease (CJD). Other diseases and conditions related to expression in the placenta might include developmental anomalies and fetal deficiencies, ovarian and endometrial cancers, reproductive disfunction and pre-natal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous and reproductive systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, reproductive, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, serum, plasma, urine, synovial

20  
25  
30

fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 141 as residues: Ala-16 to Leu-22. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions.

Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, expression in placenta would suggest a possible role in the treatment and diagnosis of developmental anomalies and fetal deficiencies, ovarian and endometrial cancers, reproductive dysfunction and pre-natal disorders.

Similarly, expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus, this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may

show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:28 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1992 of SEQ ID NO:28, b is an integer of 15 to 2006, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:28, and where b is greater than or equal to a + 14.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 19**

The translation product of this gene shares sequence homology with the murine transforming protein (See, e.g., Genbank Accession No. gil53529|embl|CAA36859.11; all references available through this accession are hereby incorporated by reference herein).

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: PIEGTPAGTGPEFPGRPTRPQRMRS LISSHPCQHLLLLLLLLFLILAILVDVKWYLVLFICISLMTSDVEHLFMCLLAIRISSWRNVY (SEQ ID NO: 290). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in activated and basal T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immunodeficiency, tumor necrosis, infection, lymphomas, auto-immunities, cancer, metastasis, wound healing, inflammation, anemias (leukemia) and other hematopoietic disorders. Similarly, polypeptides and antibodies directed to these

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in activated T-cells and the homology to a murine transforming protein indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of

various blood lineages, and in the differentiation and/or proliferation of various cell types.

The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents  
5 that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human  
10 immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g. for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating  
15 infections, tumours); haemostatic or thrombolytic activity (e.g. for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy

20 procedures. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

25 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:29 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is  
30 cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3056 of SEQ ID NO:29, b is an

integer of 15 to 3070, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:29, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 20

5       The translation product of this gene was shown to have homology to the Mus musculus ALG-2 protein, which is known to code for a Ca(2+)-binding protein required for T cell receptor-, Fas-, and glucocorticoid-induced cell death. ALG-2 mediate Ca(2+)-regulated signals along the death pathway and may play a role in the onset of Alzheimer's Disease (See e.g., Genbank Accession No.gil1213520; all  
10   references available through this accession are hereby incorporated by reference herein).

      Preferred polypeptides comprise the following amino acid sequence: NWVPT CLCPSAPCSFHLLSRFKCLFSPQRLTDIFRRYDTDQDGWIQVSYEQYLSMVFS  
15   IV (SEQ ID NO: 291), and/or QRLTDIFRRYDTDQDGWIQVSYEQYLSMVFSIV (SEQ ID NO: 292). Polynucleotides encoding these polypeptides are also provided.

      When tested against K562 cell lines, supernatants removed from cells containing this gene activated the ISRE (interferon-sensitive responsive element). Thus, it is likely that this gene activates immune or leukemia cells through the Jaks-STAT signal transduction pathway. ISRE is a promoter element found upstream in  
20   many genes which are involved in the Jaks-STAT pathway. The Jaks-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

25       A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MFYKLTLILCELSVAGVTQAASQRPLQRLPRHICSQR XPPGRCLLKAXLQTTWXXPDKPI PRLSPPLXSDPKR (SEQ ID NO: 293).

      Polynucleotides encoding these polypeptides are also provided.

      The gene encoding the disclosed cDNA is believed to reside on chromosome  
30   5. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 5.



This gene is expressed primarily in placenta, and to a lesser extent, in a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental anomalies, fetal deficiencies ovarian and endometrial cancers, reproductive dysfunction and pre-natal disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, developmental, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 143 as residues: Arg-24 to Arg-31, Ile-33 to Gly-41.

Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in placenta indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment, prevention and/or diagnosis of developmental anomalies, fetal deficiencies, ovarian and endometrial cancers, reproductive dysfunction and pre-natal disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells combined with the observed ISRE activity, and homology to the apoptosis linked, ALG-2 indicates that this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:30 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2213 of SEQ ID NO:30, b is an integer of 15 to 2227, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:30, and where b is greater than or equal to a + 14.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 21**

The translation product of this gene was shown to have homology to the human histo-blood group A transferase (See, e.g., Genbank Accession No. gblAAD26573.1|AF134413\_1 (AF134413); all references available through this  
 5 accession are hereby incorporated by reference herein) which is known to represent one of the major allogeneic antigens in both erythrocytes and tissues of humans. It has been proposed that the A phenotype is associated with the glycosyltransferase that converts the H substance associated with the O phenotype to A through the addition of alpha1-3-N-acetylgalactosamine or alpha1-3-galactosyl residues to the H antigen  
 10 Fuc-alpha1-2Gal- beta1-R. Therefore, the primary product of the histo-blood group A is its respective glycosyltransferase. Preferred polypeptides of the invention comprise the following amino acid sequence: TSSPVFSFCSMAVREPDHLQ  
 RVSLPRYNVSASLQWLPCHRIVLQPWHMCAMWELGQVLFHPVAPREGAAPS  
 PVSTLTWPSSCSHSESTMELELQF (SEQ ID NO: 294), LPCHRIV (SEQ ID NO:  
 15 296), SLQWLPCHRIVLQPW (SEQ ID NO: 297), and/or MAVREPDHLQRVSLPR (SEQ ID NO: 295). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in 12-week-old human embryo.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
 20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental anomalies, fetal deficiencies, pre-natal disorders, hematopoietic disorders, or cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential  
 25 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing fetus, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., hematopoietic, lymph, developing, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, amniotic fluid, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder,  
 30 relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in 12 week old embryo indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of developmental anomalies, fetal deficiencies, pre-natal disorders and cancer. Similarly, expression within embryonic tissue and other cellular sources marked by  
5 proliferating cells indicates that this protein may play a role in the regulation of cellular division. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus, this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Alternatively, the tissue distribution and homology to human blood group A  
10 and B glycosyltransferase enzymes indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. Representative uses are described in the "Immune Activity" and "infectious  
15 disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia.

The gene product may also be involved in lymphopoiesis, therefore, it can be  
20 used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to  
25 isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly  
30 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:31 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically

excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1274 of SEQ ID NO:31, b is an integer of 15 to 1288, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:31, and where b is greater than or equal to a + 14.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 22

The translation product of this gene shares sequence homology with CD97 (EMR1), which is thought to be important in both adhesion and signaling processes early after leukocyte activation (See, e.g., Genbank Accession No. gil784994; all references available through this accession are hereby incorporated by reference herein). EMR1 belongs to a novel family of G-protein receptors that has recently been recognized on the basis of homologous primary amino acid sequences, comprises receptors to hormones of the secretin/vasoactive intestinal peptide/glucagon family, parathyroid hormone and parathyroid hormone-related peptides, growth hormone-releasing factor, corticotropin-releasing factor, and calcitonin. Proteins with seven transmembrane segments (7TM) define a superfamily of receptors (7TMreceptors) sharing the same topology: an extracellular N-terminus, three extramembranous loops on either side of the plasma membrane, and a cytoplasmic C-terminal tail. Upon ligand binding, cytoplasmic portions of the activated receptor interact with heterotrimeric G-coupled proteins to induce various second messengers, which subsequently activate various signal transduction pathways depending upon the specific G-coupled protein associated with the receptor. Preferred polypeptides of the invention comprise the following amino acid sequence: CFKRKPKREHCSCP  
ITYQSLGDILNASFFSKRKGMQEVKLNSYVVS GTIGLKEKISLSEPVFLTFRHN  
QPGDKRTKHICVYWEGSEGGRWSTEGCSHVHSNGSYTKCKCFHLSSFAVLV  
ALAPKEDPVLTVITQVGLTISLLCLFLAILTFLLCRPIQNTSTSLHLELSLCLFLA  
HLLFLTGINRTEPEVLCSIIAGLLHFLYLACFTWMLLEGLHLFLTVRNLKVAN  
YTSTGRFKKRFMYPVGYGIPAVIIAVSAIVGPQNYGTFTHCWLKLDKGFIWSF  
MGPVAVIILINLVFYFQVLWILRSKLSSLNKEVSTIQDTRVMTFKAISQLFILGC  
SWGLGFFMVEEVGKTIGSIIAYSFTIINTLQGVLLFVVHCLLNQRQVRMEYKKW

FSGMRKGVETESTEMSRSTTQTKTEEVGKSSEIFHKGGTASSSAESTKQPQPQ  
 VHLVSAAWLKMN (SEQ ID NO: 298), and/or FFWKENLRRNGSREDFARRATQ  
 LIQSVELSIWNASFASPGKGQISEFDIVYETKRCNETRENAFLEAGNNTMDINC  
 ADALKGNLRESTAVALSLINLLGIF SEQ ID NO: 299. Polynucleotides

5 encoding these polypeptides are also provided.

Included in this invention as preferred domains are two EGF-like protein domains, which were identified using the ProSite analysis tool (Swiss Institute of Bioinformatics). First, a sequence of about forty amino-acid residues long found in the sequence of epidermal growth factor (EGF) has been shown to be present in a  
 10 large number of membrane-bound and extracellular, mostly animal proteins. Many of these proteins require calcium for their biological function and a calcium-binding site has been found to be located at the N-terminus of some EGF-like domains. Calcium-binding is crucial for numerous protein-protein interactions. We have used the N-terminal part of the EGF domain as a consensus pattern. It includes the negative N-  
 15 terminus and the possible hydroxylation site. The consensus pattern is as follows: [DEQN].[DEQN]{2}C.{3,14}C.{3,7}C.[DN].{4}[FY].C [The four C's are involved in disulfide bonds].

Preferred polypeptides of the invention comprise the following amino acid sequence: DINECETGLAKCKYKAYCRNKVGGYIC (SEQ ID NO: 300).

20 Polynucleotides encoding these polypeptides are also provided. Secondly, post-translational hydroxylation of aspartic acid or asparagine to form erythro-beta-hydroxyaspartic acid or erythro-beta-hydroxyasparagine has been identified in a number of proteins with domains homologous to (EGF). Based on sequence comparisons of the EGF-homology region that contains hydroxylated Asp or Asn, a  
 25 consensus sequence located in the N-terminal of EGF-like domains has been identified that seems to be required by the hydroxylase(s). The consensus sequence is as follows: C.[DN].{4}[FY].C.C.

Preferred polypeptides of the invention comprise the following amino acid sequence: CRNKVGGYICSC (SEQ ID NO: 301). Polynucleotides encoding these  
 30 polypeptides are also provided.

Further preferred are polypeptides comprising the calcium-binding EGF-like domain and aspartic acid and asparagine hydroxylation site listed above, and at least

5, 10, 15, 20, 25, 30, 50, or 75 additional contiguous amino acid residues of the sequence referenced in Table I for this gene and the embodiments listed herein. The additional contiguous amino acid residues is N-terminal or C-terminal to one or both of the listed domains. Alternatively, the additional contiguous amino acid residues is both N-terminal and C-terminal to one or both of the listed domains, wherein the total N- and C-terminal contiguous amino acid residues equal the specified number. The above preferred polypeptide domains are characteristic of a signature specific to EGF like proteins. Based on the sequence similarity and conserved domains, The translation product of this gene is expected to share at least some biological activities with EGF-like proteins. Such activities are known in the art, some of which are described elsewhere herein.

This gene is expressed primarily in eosinophils.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hematopoietic disorders or anemias and leukemias, immunodeficiencies, infection, lymphomas, auto-immunities and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 145 as residues: Ser-22 to Ser-30, Pro-33 to Cys-48, Asp-50 to Lys-67, Pro-117 to Ser-130. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in eosinophils combined with its homology to a known human seven transmembrane domain protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders, particularly considering the fact that the majority of 7 transmembrane receptors are tightly associated with signal transduction pathways which are integral to the modulation of the cell cycle. As such, the protein product of this gene may play a role in the regulation of cellular division, where loss of regulation may result in proliferating cells and the onset of tumors or cancer. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. In such an event, this gene is useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Similarly, the tissue distribution and homology to CD97 indicates that the protein product of this gene might be a marker for differentiation and activation of eosinophils, and therefore is useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immunities, immunodeficiencies (e.g., AIDS), immuno-suppressive conditions (transplantation) and hematopoietic disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. In addition this gene product is applicable in conditions of general microbial infection, inflammation or cancer. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are



related to SEQ ID NO:32 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
 5 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3266 of SEQ ID NO:32, b is an integer of 15 to 3280, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:32, and where b is greater than or equal to a + 14.

## 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 23

The translation product of this gene has been found to have homology to the rat neural F box protein NFB42, in addition to a conserved *Caenorhabditis elegans* C14B1.3 protein (See, e.g., Genbank Accession Nos. gil3851648|gb|AAC97505.11 (AF098301) and gil558270; all references available through these accessions are  
 15 hereby incorporated by reference herein). Preferred polypeptides of the invention comprise the following amino acid sequence: ALCPHPLILNVTVSPAPSCRHVK KVVASPSPTTMIAMDAPHSKAALDSINELPENILLELFTHVPARQLLLNCRL VCSLWRDLIDLMTLWKRKCLREGFITKDWDQPVADWKIFYFLRSLHRNLLR NPCAEEDMFAWQIDFNGGDRWKVESLPGAHGTDFPDPKVKKYFVTSYEMCL  
 20 KSQVLVDLVAEGYWEELLDTRFPDIVVKDWFAARADCGCTYQLKVQLASA DYFVLASFEPPTVTIQQWNNATWTEVSYTFSDYPRGVRYILFQHGGRTQY WAGWYGPRVTNSSIVVSPKMTRNQASSEAPGQKHGQEEAAQSPYRAVV QIF (SEQ ID NO: 302). Polynucleotides encoding these polypeptides are also provided.

25 The gene encoding the disclosed cDNA is believed to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in immune cells, especially primary dendritic cells, and T cells, and to a lesser extent in a variety of other tissues including breast,  
 30 keratinocytes, epididymus (cauda), lung, multiple sclerosis, endometrial stromal cells, IL4 induced umbilical vein endothelial cells, fetal kidney, fetal dura mater, rejected kidney, and osteoblasts.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer and other proliferative disorders, particularly of the immune system or endothelial cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 146 as residues: Pro-41 to Cys-47, Phe-52 to Gly-59, Pro-62 to His-70. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in immune cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product in T-cells indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g., by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities,

such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types.

The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions and as nutritional supplements. It may also have a very wide range of biological activities although no evidence for any is provided in the specification. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g., for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g., for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g., for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumours); haemostatic or thrombolytic activity (e.g., for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g., for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:33 and may have been publicly available prior to conception of

the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general  
5 formula of a-b, where a is any integer between 1 to 1283 of SEQ ID NO:33, b is an integer of 15 to 1297, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:33, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 24

10 The translation product of this gene shares sequence homology with the human, mouse, and bovine dopamine hydroxylase which is thought to be important in the modification of dopamine, a neurotransmitter (See Genbank Accession Nos. gil30474, gil162965, and/or gil2358082; all references available through these accessions are hereby incorporated by reference herein). Preferred polypeptides of the  
15 invention comprise the following amino acid sequence: RQRSWNPGT  
NCYHPNMPDAFLT CETVIFAWAIGGEGFSYPPHVGLSLGTPLDPHYVLLEVH  
YDNPTYEEGLIDNSGLRFLFYTMDIRKYDAGVIEAGLWVSLFHTIPPGMPEF  
QSEGHCTLECLEEALEAEKPSGIHVFAVLLHAHLAAGR GIRLRHFRKGKEMKL  
LAYDDDFDFNFQEFQYLKEEQTILPGDNLITECRYNTKDRAEMTWGGLSTR  
20 SEMCLSYLLYYPRINLTRCASIPDIMEQLQFIGVKEIYRPVTTWPFIKSPKQYK  
NLSFMDAMNKF KWKKEGLSFNKLVLSPVNVRC SKTDNAEWSIPRND SIT  
SRYR KTL (SEQ ID NO: 303). Polynucleotides encoding these polypeptides are also provided.

A preferred polypeptide fragment of the invention comprises the following  
25 amino acid sequence: MCCWPLLLLWGLLPGTAAGGSGRTYPHRTLLDSE GK  
YWLGWSQRGSQIAFRLQVRTAGYVGFGFSPTGAMASADIVVGGVAHGR  
PYLQDYFTNANRELKKDAQQDYHLEYAMENSTHTIIEFTRELHTCDINDKS  
ITDSTVRVIWAYHHE DAGEAGPKYHDSNRGTKSLRLLNPEKTSVLSTALPYF  
DLVNQDVPINPKDTTYWCQMFKIPVFQEKHHVIKVEPVIQRGHESLVHHILL  
30 YQCSNNFNDSPGIRARIAITPTCPMHSSPV KL (SEQ ID NO: 304).

Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in brain, the pulmonary system, and to a lesser extent in kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurological and behavioral disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, endocrine, or cancerous and wounded tissues) or bodily fluids (e.g., sputum, lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 147 as residues: Ser-33 to Trp-38, Gly-40 to Gly-45, Asn-93 to Asp-105, Thr-128 to Thr-137, Glu-150 to Lys-167, Pro-197 to Tyr-203, Cys-242 to Asn-247, Ser-253 to Tyr-258, His-307 to Glu-314, Glu-357 to Gly-362, Trp-373 to Gln-378, Ser-402 to Glu-408. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain and homology to a protein involved in the modification of dopamine indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive

compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system. Alternatively, the homology to dopamine hydroxylase indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers, particularly Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g., diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g., hyper-, hypothyroidism), parathyroid (e.g., hyper-, hypoparathyroidism), hypothalamus, and testes. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:34 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2170 of SEQ ID NO:34, b is an integer of 15 to 2184, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:34, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 25

When tested against Jurkat T-cell lines, supernatants removed from cells containing this gene activated the gamma activating sequence (GAS), a promoter element found upstream of many genes which are involved in the Jak-STAT pathway.

The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells. Thus, it is likely that this gene activates T-cells through the JakStat signal transduction pathway.

This gene is expressed in a variety of human normal and diseased tissues including breast, infant adrenal gland, skin tumor, colon, pituitary, Wilm's tumor, and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, breast cancer and other proliferative disorders, afflicting endocrine or endothelial tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system or of breast and/or breast lymph nodes, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, endocrine, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in breast, infant adrenal gland, skin tumor, colon, pituitary, and Wilm's tumor, and biological activity in activating the GAS promoter element indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers, particularly Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g., diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g., hyper-, hypothyroidism), parathyroid (e.g., hyper-, hypoparathyroidism), hypothalamus, and testes. Alternatively, the tissue distribution and biological activity indicates that polynucleotides and polypeptides corresponding to this gene are useful for the

diagnosis and treatment of cancer and other proliferative disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells, (i.e., breast, skin and Wilm's tumors) indicates that this protein may play a role in the regulation of cellular division. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. In such an event, this gene is useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:35 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 935 of SEQ ID NO:35, b is an integer of 15 to 949, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:35, and where b is greater than or equal to a + 14.



**FEATURES OF PROTEIN ENCODED BY GENE NO: 26**

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the  
5 following amino acid sequence: TGTFWSPRSQRRGCCGRRAPRPEAMENGAVYS  
PTTEEDPGPARGPRSGLAAYFFMGRLLPLRRVLKGLQLLLSLLAFICEEVVSQ  
CTLCGGLYFFEFVSCSAFLLSLLILIVYCTPFYERVDTTKVKSSDFYITLGTGCV  
FLLASIIFVSTHDRTSAEIAAIVFGFIASFMLLDFFITMLYEKRQESQLRKPENTT  
RAEALTEPLNA (SEQ ID NO: 305). Polynucleotides encoding these polypeptides  
10 are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 3. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 3.

This gene is expressed primarily in dendritic cells, and to a lesser extent in  
15 melanocytes, fetal liver and spleen and several other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, inflammation, and disorders of the hepatic and immune systems.

20 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., hematopoietic,  
25 hepatic, immune, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 149 as residues: Phe-63 to Ser-75, Thr-97 to Ser-102, Glu-128 to Arg-143. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product in dendritic cells indicates a role in the regulation of the proliferation; survival; differentiation; and/or activation of potentially all hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes that may also suggest a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also used as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, scleroderma and tissues. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Alternatively, the tissue distribution in fetal liver indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of liver disorders and cancers (e.g., hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells). In addition the expression in fetus would suggest a useful role for the protein product in developmental abnormalities, fetal deficiencies, pre-natal disorders and various wound-healing models and/or tissue trauma. Furthermore, the protein may also be used to determine biological activity, raise

antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

- 5 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:36 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is
- 10 cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3324 of SEQ ID NO:36, b is an integer of 15 to 3338, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:36, and where b is greater than or equal to a + 14.

15

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 27

In a specific embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the

- 20 following amino acid sequence: ASAPRVMRGHLAGFPALSGLASVCLWATFSA  
QLPGPVAATSWTPAPLGCSAARSGPEKRLGTAAPGSAASLAQAGPGAPCRV  
LPVDPAPAALNVREPGWLGGFLFDGALLQVLLNFLRKSTDVLMDTREAESLEV  
E (SEQ ID NO: 306).

- 25 In another embodiment polypeptides of the invention comprise the following amino acid sequence: NKLHSFPVFLSQLLLDRQLLHAPQTLTPHCGGSSRPGP  
SHPPWLLIQLPCVHVALWQMLRDFSDSRITPSTLTTQPAAQTAAPAKDQES  
DIVGGEGILCDIAFLQEDHPLGVGGASAPSSRRELSRRGVHTQTLPEDGTLHG  
TPSSSFDCGIKYIISWPLAPGCDLPSLELSLVCKGVSSCMGFAAG (SEQ ID NO:  
307). Polynucleotides encoding these polypeptides are also provided.

- 30 This gene is expressed primarily in endothelial cells, lung, and fetal kidney, and to a lesser extent in epididymis, keratinocytes and cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cardiovascular diseases involving endothelial cell disturbances such as atherosclerosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cardiovascular, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 150 as residues: Arg-47 to Leu-54. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in endothelial cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating disorders of endothelial cells such as atherosclerosis, vasculitis, cardiovascular disease, and emphysema. The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions and as nutritional supplements. The polypeptide may possess a wide range of undetected biological activities. Typical of these are cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g., for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of haematopoiesis (e.g., for treating anaemia or as adjunct to chemotherapy); stimulation of growth of bone, cartilage, tendons, ligaments and/or nerves (e.g., for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g., for treating infections, tumours); haemostatic or

thrombolytic activity (e.g., for treating haemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g., for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative disease; for regulation of metabolism, behaviour, and many others. Also contemplated is the use of the  
5 corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are  
10 related to SEQ ID NO:37 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
15 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1549 of SEQ ID NO:37, b is an integer of 15 to 1563, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:37, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 28

20 In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: PGRPTRPTKNKVCVCLGMLFWAYPICVFIDSL  
25 SCQPCLWSTGATSHFNSPTTSPLFTLFMPCALAPNPFT QLGKLDDR (SEQ ID NO: 308). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in meningima.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
30 not limited to, tumors or disorders of the central nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For

a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, or cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 151 as residues: His-29 to Thr-34. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in meningioma indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception, as well as disorders of the meninges such as meningioma and meningitis. In addition, the gene or gene product may also play a role in the treatment and/or detection of developmental disorders associated with the developing embryo, sexually-linked disorders, or disorders of the cardiovascular system. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:38 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1034 of SEQ ID NO:38, b is an integer of 15 to 1048, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:38, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 29

The translation product of this gene has been shown to encode a human brain specific mitochondrial carrier (Genbank Accession No. gi3851540|gb|AAD04346.11 (AF078544); all references available through this accession are hereby incorporated herein by reference) which shares sequence homology with the human body weight disorder associated gene C5 product which is known to be differentially expressed in obese compared to lean mice (See GeneSeq Accession No. R91281). Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with mitochondrial carriers proteins. Such activities are known in the art, some of which are described in Sanchis et al, J. Biol. Chem. 273:34611-34615 (1998), incorporated herein by reference.

Included in this invention as preferred domains are mitochondrial energy transfer protein (METP) domains, which were identified using the ProSite analysis tool (Swiss Institute of Bioinformatics). Structurally, members of the family of mitochondrial energy transfer proteins consist of three tandem repeats of a domain of approximately one hundred residues. Each of these domains contains two transmembrane regions. As a signature pattern, we selected one of the most conserved regions in the repeated domain, located just after the first transmembrane region. To detect this widespread family of proteins, a consensus sequence was developed that contains the most conserved regions in the repeated domain. The consensus pattern is as follows: P.[DE].[LIVAT][RK].[LRH][LIVMFY][QMAIGV].

Preferred polypeptides of the invention comprise the following amino acid sequences: PVDLTKTRLQ (SEQ ID NO: 309) and PTDVLKIRMQ (SEQ ID NO: 310). Polynucleotides encoding these polypeptides are also provided.

Further preferred are polypeptides comprising the METP domains of the  
5 sequence listed above, and at least 5, 10, 15, 20, 25, 30, 50, or 75 additional  
contiguous amino acid residues of the sequence referenced in Table I for this gene.  
The additional contiguous amino acid residues is N-terminal or C-terminal to the  
METP domain. Alternatively, the additional contiguous amino acid residues is both  
N-terminal and C-terminal to the METP domain, wherein the total N- and C-terminal  
10 contiguous amino acid residues equal the specified number. The above preferred  
polypeptide domain is characteristic of a signature specific to mitochondrial energy  
transfer proteins.

The gene encoding the disclosed cDNA is believed to reside on chromosome  
X. Accordingly, polynucleotides related to this invention are useful as a marker in  
15 linkage analysis for chromosome X.

This gene is expressed primarily in brain, and to a lesser extent, in T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions which include, but are  
20 not limited to, neurological and behavioral disorders and immune disorders and/or  
obesity. Similarly, polypeptides and antibodies directed to these polypeptides are  
useful in providing immunological probes for differential identification of the  
tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,  
particularly of the digestive, immune, and nervous systems, expression of this gene  
25 at significantly higher or lower levels is routinely detected in certain tissues or cell  
types (e.g., immune, neural, or cancerous and wounded tissues) or bodily fluids (e.g.,  
lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell  
sample taken from an individual having such a disorder, relative to the standard gene  
expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
30 individual not having the disorder.



Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 152 as residues: Gln-189 to Gly-195. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain and homology to mitochondrial carrier proteins indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:39 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

formula of a-b, where a is any integer between 1 to 1416 of SEQ ID NO:39, b is an integer of 15 to 1430, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:39, and where b is greater than or equal to a + 14.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 30

Preferred polypeptides of the invention comprise the following amino acid sequence: MTFGSTISPTSTHASPGLGCCSWLLEDLEEQLYCSAFEEAALTR  
RICNPTSCWLPLDMELLHRQVLALQTQRVLLGMWLRRAWDTWVSPRRVAP  
GSRCLLTASHPCTEKRRKASAXQRNLGYPLAMLCLLVLTGLSVLIVAIHILEL  
10 LIDEAAMPRGMQGTSLGQVSFSKLGSFGAVIQVVLIFYLMVSSVVGIFYSSPLF  
RSLRPRWHDAMTQIIGNCVCLLVSSALPVFSRTLGLTRFDLLGDFGRFNWL  
GNFYIVFLYNAAFAGLTTLCLVKTFTA AVRAELIRAFGLDRLPLPVSGFPQAS  
RKTQHQ (SEQ ID NO: 311). Polynucleotides encoding such polypeptides are also provided.

15 This gene is expressed primarily in immune system tissues (e.g. resting T-cells, primary dendritic cells, and neutrophils, apoptotic T-cells) and umbilical vein. This gene is expressed to a lesser extent in the gastrointestinal tissue (e.g. small intestine, colon), brain (e.g. cerebellum, frontal cortex), aorta endothelial cells, skin tumor, embryonic tissue, thymus, and cancers (e.g. cheek, breast, synovial).

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for  
25 differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and gastrointestinal tract expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, gastrointestinal, cancerous and wounded tissues) or bodily fluids (e.g., amniotic, serum, plasma, urine, synovial fluid and  
30 spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 153 as residues: Asp-21 to Ser-29. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in immune cells (e.g. T-cells, dendritic cells, 5 neutrophils) indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the 10 proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene 15 product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and 20 tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits 25 hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The tissue distribution in skin tumors and cancerous tissue (e.g. cheek, breast, synovial sarcoma) indicates that polynucleotides and polypeptides corresponding to this gene are useful for the 30 diagnosis and treatment of cancer and other proliferative disorders. Expression in cellular sources such as embryonic tissue marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Additionally, the

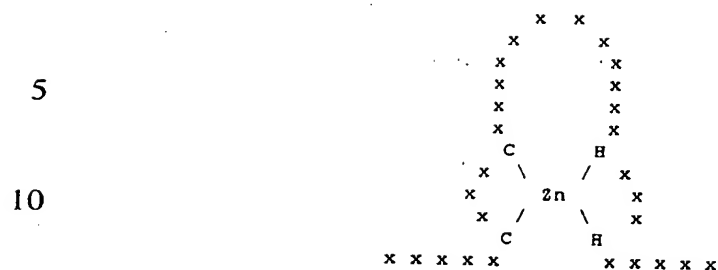
expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. In such an event, this gene is useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from  
5 early hematopoietic stem and committed progenitor cells. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or  
10 immunotherapy targets for the above listed tissues. The tissue distribution in cerebellum and frontal cortex indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative  
15 Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction,  
20 aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.  
25 Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional  
30 supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:40 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2089 of SEQ ID NO:40, b is an integer of 15 to 2103, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:40, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 31

The polypeptide of this gene has been determined to have a zinc finger ( Zinc finger, C2H2 type) domain at about amino acid position 16-50 of the amino acid sequence referenced in Table 1 for this gene. Therefore,

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: LCVCLVYLCMYGVCLCVIVCVSGVSLCLYVWGV SVC DCVSVFMCVCLCVIFCVYVGKPRTEHYHSPHLAKQKAFREMCGRHDVSAAGIF QSYV (SEQ ID NO: 312). Polynucleotides encoding these polypeptides are also provided. 'Zinc finger' domains are nucleic acid-binding protein structures first identified in the Xenopus transcription factor TFIIA. These domains have since been found in numerous nucleic acid-binding proteins. A zinc finger domain is composed of 25 to 30 amino-acid residues. There are two cysteine or histidine residues at both extremities of the domain, which are involved in the tetrahedral coordination of a zinc atom. It has been proposed that such a domain interacts with about five nucleotides. A schematic representation of a zinc finger domain is shown below:



Many classes of zinc fingers are characterized according to the number and positions of the histidine and cysteine residues involved in the zinc atom coordination. In the first class to be characterized, called C2H2, the first pair of zinc coordinating residues are cysteines, while the second pair are histidines. A number of experimental reports have demonstrated the zinc- dependent DNA or RNA binding property of some members of this class. Some of the proteins known to include C2H2-type zinc fingers are listed below. We have indicated, between brackets, the number of zinc finger regions found in each of these proteins; a '+' symbol indicates that only partial sequence data is available and that additional finger domains is present.

This gene is expressed primarily in salivary gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, salivary gland related diseases, diseases of the mouth, and other digestive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., saliva, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 154 as residues: Gly-46 to His-54. Polynucleotides encoding said polypeptides are also provided.

5 The tissue distribution indicates that the protein products of this gene are useful for diagnosis and treatment of salivary gland related diseases (mumps, calculi formation in ducts, sarcoidosis, facial palsy, tumors, Sjogrens Syndrome) and other digestive system disorders. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a  
10 nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are  
15 related to SEQ ID NO:41 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general  
20 formula of a-b, where a is any integer between 1 to 2335 of SEQ ID NO:41, b is an integer of 15 to 2349, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:41, and where b is greater than or equal to a + 14.

### **FEATURES OF PROTEIN ENCODED BY GENE NO: 32**

25 This gene is expressed primarily in fetal tissue (e.g. spleen, liver, brain), cancerous tissues (e.g. ovarian, colon, stomach, parathyroid) and to a lesser extent in immune cells and tissue (e.g. B-cells, T-cells, bone marrow), and reproductive organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
30 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, particularly of the colon and ovaries, disorders of the developing fetus, neurodegenerative conditions, and immune system disorders.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is  
5 routinely detected in certain tissues or cell types (e.g., immune, reproductive, neural, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 155 as residues: Lys-35 to Lys-47. Polynucleotides encoding said polypeptides are also provided.

The expression of this gene within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular  
15 division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern  
20 formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of  
25 potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types  
30 of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in



modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. The tissue distribution in immune cells (such as T-cells and B-cells) and immune tissues (bone marrow)

5 indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation;  
10 survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene  
15 product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and  
20 tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits  
25 hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The tissue distribution in parathyroid indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine  
30 disorders and cancers. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", and "Binding Activity" sections below, in Example 11, 17, 18, 19, 20 and 27, and elsewhere herein. Briefly, the protein can be used for

the detection, treatment, and/or prevention of Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g. diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g. hyper-, hypothyroidism), parathyroid (e.g. hyper-, hypoparathyroidism), hypothalamus, and testes. Additionally, the tissue distribution in brain tissue indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:42 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or

more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1545 of SEQ ID NO:42, b is an integer of 15 to 1559, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:42, and where b is greater than or equal to a + 14.

5

### FEATURES OF PROTEIN ENCODED BY GENE NO: 33

When tested against U937 Myeloid cell lines, supernatants removed from cells containing this gene activated the GAS assay. Thus, it is likely that this gene activates myeloid cells through the Jak-STAT signal transduction pathway. The gamma  
10 activating sequence (GAS) is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and  
15 differentiation of cells.

This gene is expressed primarily in skin tumors.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
20 not limited to, skin disorders, particularly skin cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell  
25 types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 156 as residues: Pro-38 to Gly-44, Phe-56 to Thr-64. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in skin indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment, diagnosis, and/or prevention of various skin disorders including congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. 5 keratoses, Bowen's Disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's Disease, mycosis fungoides, and Kaposi's sarcoma), injuries and inflammation of the skin (i.e. wounds, rashes, prickly heat disorder, psoriasis, dermatitis), atherosclerosis, urticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, 10 pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. Moreover, such disorders may predispose increased susceptibility to viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, athlete's foot, and ringworm). Protein, as well as, antibodies directed against the 15 protein may show utility as a tumor marker and immunotherapy targets for the above listed tumors and tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:43 and may have been publicly available prior to conception of 20 the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1752 of SEQ ID NO:43, b is an 25 integer of 15 to 1766, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:43, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 34

The translation product of this gene shares sequence homology with mitogen- 30 induced prostate carcinoma (mouse) which is thought to be important in the etiology of cancer. In this respect, this gene is mitogen-induced and/or involved in cell proliferation.

Preferred polypeptides of the invention comprise the following amino acid sequence: GHMPYGWLTEIRAVYPAFDKNNPSNKLVSTSNTVTAHIKKF  
TFVCMALSLTLCFVMFWTPNVSEKILIDIIGVDFAFELCVVPLRIFSFFPVPVT  
VRAHLTGWLMTLKKTFVLAPSSVLRIIVLIASLVVLPYLGVBHGATLGVGSLLA  
5 GFVGESTMVAIAACYVYRKQKKK MENESATEGEDSAMTDMPTTEEVTDIVE  
MREENE (SEQ ID NO: 313) and/or QVVFVAILLHSHLECREPLLIPILSLYMGA  
LVRCTTLCLGYKNIHDIIPDRSGPELGGDATIRKMLSFWWPLALILATQRI  
PIVNLFVSRDLGGSSAATEAVAILTATYPV (SEQ ID NO: 314). Polynucleotides  
encoding these polypeptides are also provided.

10 The gene encoding the disclosed cDNA is believed to reside on chromosome  
5. Accordingly, polynucleotides related to this invention are useful as a marker in  
linkage analysis for chromosome 5.

This gene is expressed primarily in early infant and adult brain, retina, fetal  
tissue (e.g., liver, spleen, whole embryo) and to a lesser extent in immune cells (e.g.,  
15 monocytes and T-cells), colon, and parathyroid tumor tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions which include, but are  
not limited to, cancers, disorders of the immune system and nervous system.

20 Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
providing immunological probes for differential identification of the tissue(s) or cell  
type(s). For a number of disorders of the above tissues or cells, particularly of the  
metabolic system (cancers), expression of this gene at significantly higher or lower  
levels is routinely detected in certain tissues or cell types (e.g., immune, neural,  
25 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial  
fluid and spinal fluid) or another tissue or cell sample taken from an individual having  
such a disorder, relative to the standard gene expression level, i.e., the expression  
level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic  
30 epitopes shown in SEQ ID NO: 157 as residues: Arg-122 to Ser-139, Met-144 to Glu-  
149. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution and homology to mitogen induced prostate carcinoma (mouse) indicates that polynucleotides and polypeptides corresponding to this gene are useful for the study and treatment of cancers, including but not limited to the colon, parathyroid, and adrenal glands. Moreover, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. The tissue distribution in immune cells (T-cells, monocytes) indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or

activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

- 5            Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, 10 such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that 15 influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are 20 useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, 25 Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, 30 and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to  
5 identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are  
10 related to SEQ ID NO:44 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
15 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2558 of SEQ ID NO:44, b is an integer of 15 to 2572, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:44, and where b is greater than or equal to a + 14.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 35**

20 This gene is expressed primarily in adult pulmonary tissue, umbilical vein, prostate, and fetal tissue (e.g., heart).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
25 not limited to, diseases of the pulmonary system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary system, expression of this gene at significantly higher or lower levels is routinely detected in  
30 certain tissues or cell types (e.g., pulmonary, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the



standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 158 as residues: Arg-45 to Gly-51, Glu-75 to Asn-81.

5 Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in pulmonary tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of disorders associated with developing lungs, particularly in premature infants where the lungs are the last tissues to develop. Additionally, the tissue distribution indicates  
10 that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and intervention of lung tumors, since the gene is involved in the regulation of cell division, particularly since it is expressed in fetal tissue. Moreover, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may  
15 show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

20 Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have  
25 applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue  
30 differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in

proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:45 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 512 of SEQ ID NO:45, b is an integer of 15 to 526, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:45, and where b is greater than or equal to a + 14.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 36**

This gene is expressed primarily in adipose tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, fat metabolism. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 159 as residues: Pro-96 to Ser-106. Polynucleotides  
5 encoding said polypeptides are also provided.

The tissue distribution in adipose tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of obesity and other metabolic and endocrine conditions or disorders. Furthermore, the protein product of this gene may show utility in ameliorating conditions which occur  
10 secondary to aberrant fatty-acid metabolism (e.g. aberrant myelin sheath development), either directly or indirectly. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly  
15 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:46 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
20 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1018 of SEQ ID NO:46, b is an integer of 15 to 1032, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:46, and where b is greater than or equal to a + 14.

## 25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 37**

This gene is expressed primarily in adult brain tissue, testes, placenta, kidney, infant and fetal tissue (e.g., liver, spleen, lung) and to a lesser extent in immune cells (e.g., T-cells and neutrophils) and in cancerous tissues (e.g., ovarian tumor, Hodgekins lymphoma, pancreas, T-cell).

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are

not limited to, CNS disorders, disorders of the testicles, cancer, particularly ovarian, pancreatic, T-cell, and Hodgekin's lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, CNS, and testes expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, urogenital, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Moreover, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders"

and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. The tissue distribution indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities,

such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Additionally, the tissue distribution in testes indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:47 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or

more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2666 of SEQ ID NO:47, b is an integer of 15 to 2680, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:47, and where b is greater than or equal to a + 14.

5

### FEATURES OF PROTEIN ENCODED BY GENE NO: 38

When tested against fibroblast cell lines, supernatants removed from cells containing this gene activated the EGR1 assay. Thus, it is likely that this gene activates fibroblast cells through a signal transduction pathway. Early growth response 1 (EGR1) is a promoter associated with certain genes that induces various tissues and cell types upon activation, leading the cells to undergo differentiation and proliferation.

This gene is expressed primarily in endometrial stromal cells, endometrial tumors, keratinocytes, fetal tissue (e.g. liver, spleen) and to a lesser extent in endothelial cells and immune cells (e.g., T-cells).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, endometrial carcinoma and immune cells disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in the endometrium indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating female infertility. The protein product is likely involved in preparation of the endometrium of implantation and could be administered either topically or orally. Alternatively, this gene could be

transfected in gene-replacement treatments into the cells of the endometrium and the protein products could be produced. Similarly, these treatments could be performed during artificial insemination for the purpose of increasing the likelihood of implantation and development of a healthy embryo. In both cases this gene or its gene product could be administered at later stages of pregnancy to promote healthy development of the endometrium. Additionally, polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of endometrial carcinoma. The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of cancer and other proliferative disorders. Expression within embryonic tissue and other cellular sources marked by proliferating cells indicates that this protein may play a role in the regulation of cellular division. Additionally, the expression in hematopoietic cells and tissues indicates that this protein may play a role in the proliferation, differentiation, and/or survival of hematopoietic cell lineages. In such an event, this gene is useful in the treatment of lymphoproliferative disorders, and in the maintenance and differentiation of various hematopoietic lineages from early hematopoietic stem and committed progenitor cells. Similarly, embryonic development also involves decisions involving cell differentiation and/or apoptosis in pattern formation. Thus this protein may also be involved in apoptosis or tissue differentiation and could again be useful in cancer therapy. The tissue distribution in immune cells such as helper T-cells indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such



as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The tissue distribution in keratinocytes indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment, diagnosis, and/or prevention of various skin disorders. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", "infectious disease", and "Regeneration" sections below, in Example 11, 19, and 20, and elsewhere herein. Briefly, the protein is useful in detecting, treating, and/or preventing congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. keratoses, Bowen's Disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's Disease, mycosis fungoides, and Kaposi's sarcoma), injuries and inflammation of the skin (i.e. wounds, rashes, prickly heat disorder, psoriasis, dermatitis), atherosclerosis, urticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. In addition, such disorders may predispose increased susceptibility to viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, athlete's foot, and ringworm). Moreover, the protein product of this gene may also be useful for the treatment or diagnosis of various connective tissue disorders (i.e., arthritis, trauma, tendonitis, chondromalacia and inflammation, etc.), autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, dermatomyositis, etc.), dwarfism, spinal deformation, joint abnormalities, and

chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:48 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1716 of SEQ ID NO:48, b is an integer of 15 to 1730, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:48, and where b is greater than or equal to a + 14.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 39**

This gene is expressed primarily in LNCAP cells (prostate cell line) and retina derived N2b5HR cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, prostate cancer and eye disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male urogenital and reproductive system expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative

to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 162 as residues: Asn-50 to Ser-57. Polynucleotides  
5 encoding said polypeptides are also provided.

The expression in prostate may indicate the gene or its products can be used in the disorders of the prostate, including inflammatory disorders, such as chronic prostatitis, granulomatous prostatitis and malacoplakia, prostatic hyperplasia and prostate neoplastic disorders, including adenocarcinoma, transitional cell carcinomas,  
10 ductal carcinomas, squamous cell carcinomas, or as hormones or factors with systemic or reproductive functions. The tissue distribution in retina indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and/or detection of eye disorders including blindness, color blindness, impaired vision, short and long sightedness, retinitis pigmentosa, retinitis proliferans,  
15 and retinoblastoma, retinchoroiditis, retinopathy and retinoschisis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or  
20 immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:49 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically  
25 excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1261 of SEQ ID NO:49, b is an integer of 15 to 1275, where both a and b correspond to the positions of nucleotide  
30 residues shown in SEQ ID NO:49, and where b is greater than or equal to a + 14.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 40**

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the  
5 following amino acid sequence: RCCCRGCSCRARLCPPARSTAVAPECRGAHPSR  
AMRPGTALQAVLLAVLLVGLRAATGRLLSGQPVCRGGTQRPCYKVIYFHD  
TSRRLNFEEAKEACRRGWRPASQHRVLKMNRN (SEQ ID NO: 315).

Polynucleotides encoding these polypeptides are also provided.

A preferred polypeptide fragment of the invention comprises the following  
10 amino acid sequence: MRPGTALQAVLLAVLLVGLRAATGRLLSGQPVCRGG  
TQRPCYKVIYFHDTSRRLNFEEAKEACRRGWRPASQHRVLKMNRN (SEQ ID  
NO: 316). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in smooth muscle and human thyroid and to a  
15 lesser extent in amniotic cells and human endometrial stromal cells-treated with  
progesterone.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions which include, but are  
not limited to, thyroid disorders. Similarly, polypeptides and antibodies directed to  
20 these polypeptides are useful in providing immunological probes for differential  
identification of the tissue(s) or cell type(s). For a number of disorders of the above  
tissues or cells, particularly of the endocrine system, expression of this gene at  
significantly higher or lower levels is routinely detected in certain tissues or cell types  
(e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine,  
25 synovial fluid and spinal fluid) or another tissue or cell sample taken from an  
individual having such a disorder, relative to the standard gene expression level, i.e.,  
the expression level in healthy tissue or bodily fluid from an individual not having the  
disorder.

Preferred polypeptides of the present invention comprise immunogenic  
30 epitopes shown in SEQ ID NO: 163 as residues: Ser-75 to Leu-81. Polynucleotides  
encoding said polypeptides are also provided.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of endocrine disorders of the thyroid.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:50 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1748 of SEQ ID NO:50, b is an integer of 15 to 1762, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:50, and where b is greater than or equal to a + 14.

#### 15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 41**

This gene is expressed primarily in human testes tumor and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the testicles including but not limited to testicular cancer and immune system disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male reproductive system and immune system expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 164 as residues: His-31 to Gly-41. Polynucleotides encoding said polypeptides are also provided.

5 The tissue distribution in testes, particularly testicular tumors, indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) 10 are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as 15 hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. The tissue distribution in bone marrow indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in 20 Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the 25 treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory 30 bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity

disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:51 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2045 of SEQ ID NO:51, b is an integer of 15 to 2059, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:51, and where b is greater than or equal to a + 14.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 42**

The translation product of this gene shares sequence homology with protocadherins, which are related to cadherin, and possess cell adhesive ability. Cadherins are glycosylated integral membrane proteins that are involved in cell-cell adhesion.

This gene is expressed primarily in brain (infant, adult frontal lobe, manic depression tissue) and to a lesser extent in epididymus, healing groin wounds, ovary, adipocytes, and fetal tissue (e.g., kidney and retina).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurodegenerative disorders, impaired male and female fertility, developmental disorders, fibrosis, and manic depression. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system and reproductive system expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, reproductive, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 165 as residues: Val-35 to Lys-41, Ser-68 to Gln-73, Glu-88 to Glu-93, Arg-156 to Gly-163, Ala-199 to Gly-206, Asp-216 to Ser-226, Thr-249 to Asn-254, Asp-339 to Pro-345, Ile-370 to Gly-379, Pro-429 to Glu-434, Arg-461 to Pro-466, Ala-475 to Thr-482, Pro-585 to Gly-593, Glu-631 to Gln-639, Pro-674 to Pro-682, Gln-715 to Gly-720, Ser-736 to Arg-742. Polynucleotides encoding said polypeptides are also provided.

BLAST analysis reveals high homology to protocadherin sequences. These sequences are related to cadherin, and possess cell adhesive ability. Such proteins may have regulatory functions in the cell, as well as the cell-cell adhesive properties. Antibodies produced against these sequences are useful for modulating the binding activity of these protocadherins, and can be used therapeutically. The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease,



Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The tissue distribution in epididymus indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. Moreover, the expression within fetal tissue (e.g., kidney and retina) and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including blindness, cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent

of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:52 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3268 of SEQ ID NO:52, b is an integer of 15 to 3282, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:52, and where b is greater than or equal to a + 14.

### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 43

Preferred polypeptides of the invention comprise the following amino acid sequence:

IRHEQQGEEDDEHARPLAESLLLAIADLLFCPDFTVQSHRRSTVDSAEDVHSL  
DSCEYTWEAGVGFAHSPQPNYIHDMNRMELLKLLLTCFSEAMYLPPAPESGS  
TNPWVQFFCSTENRHALLFTSLLNTVCAYPDVGYGIPYNHLLFSDYREPLVE  
EAAQVLIVTLDHDSASSASPTVDGTTTGTAMDDADPPGPENLFVNYLSRIHRE  
5 EDFQFILKGIARLLSNPLLQTYLPNSTKKDPVPPGAASSLLEALRLQQEIPLLRA  
EEQRRPRHPCPHLLPQRCPPGRSV (SEQ ID NO: 317). Polynucleotides encoding  
such polypeptides are also provided.

This gene is expressed primarily in brain, breast, breast cancer tissue and to a lesser extent in epididymus, amniotic cells, and embryo tissue.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurodegenerative disorders, impaired CNS function, male sterility, and breast cancer. Similarly, polypeptides and antibodies directed to these  
15 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and reproductive systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, male reproductive, cancerous and wounded tissues) or  
20 bodily fluids (e.g., amniotic, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic  
25 epitopes shown in SEQ ID NO: 166 as residues: Pro-22 to Pro-31, Ser-38 to His-43, Asp-74 to Leu-79, Asp-113 to Glu-121, Leu-157 to Val-166, Ala-189 to Arg-196, Gln-206 to Arg-211. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain, particularly in the cerebellum, indicates polynucleotides and polypeptides corresponding to this gene are useful for the  
30 detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11,

15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal  
5 cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in  
10 normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The tissue distribution in epididymus indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of  
15 conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment  
20 and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few  
25 possible target indications. The expression in the breast tissue may indicate its uses in the diagnosis and/or treatment of breast neoplasia and breast cancers, such as fibroadenoma, papillary carcinoma, ductal carcinoma, Paget's Disease, medullary carcinoma, mucinous carcinoma, tubular carcinoma, secretory carcinoma and apocrine carcinoma, as well as juvenile hypertrophy and gynecomastia, mastitis and  
30 abscess, duct ectasia, fat necrosis and fibrocystic diseases. Moreover, the expression within embryonic tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may

show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions  
5 involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of  
10 potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types  
15 of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new  
20 insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or  
25 immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:53 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically  
30 excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

formula of a-b, where a is any integer between 1 to 1846 of SEQ ID NO:53, b is an integer of 15 to 1860, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:53, and where b is greater than or equal to a + 14.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 44

Contact of cells with supernatant expressing the product of this gene increases the permeability of monocytes to calcium. Thus, it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product of this gene binds a receptor on the surface of the monocyte cell. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating monocyte cells.

This gene is expressed primarily in CD34 positive cells derived from human cord blood.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hematopoietic disorders; immune dysfunction; defects in hematopoietic stem and progenitor cells; susceptibility to chemotherapy and irradiation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 167 as residues: Ala-38 to Leu-59, Ala-63 to Thr-71, Lys-82 to Leu-91, Glu-97 to Ser-107, Gln-143 to Ala-149, Ile-153 to Leu-158, Ser-169 to Arg-182. Polynucleotides encoding said polypeptides are also provided.

Elevated expression of this gene product in CD34 positive hematopoietic cells indicates that it is expressed by early stem and progenitor cells of the hematopoietic lineages. Therefore, this may represent a soluble factor that is able to control the survival, proliferation, differentiation, or activation of all hematopoietic lineages, including stem and progenitor cells. Thus, it could be quite useful, for example, in ex vivo expansion of stem cell numbers for hematopoietic disorders or for cancer patients. Alternately, it may represent a factor that influences the hematopoietic microenvironment by affecting stromal cells that release other factors required for hematopoietic development. Additionally, the tissue distribution in CD34 positive cells also indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of hematopoietic related disorders such as anemia, pancytopenia, leukopenia, thrombocytopenia or leukemia since stromal cells are important in the production of cells of hematopoietic lineages. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the uses include bone marrow cell ex-vivo culture, bone marrow transplantation, bone marrow reconstitution, radiotherapy or chemotherapy of neoplasia.

The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:54 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 756 of SEQ ID NO:54, b is an integer of 15 to 770, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:54, and where b is greater than or equal to a + 14.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 45

This gene is expressed primarily in breast and 12-week old human embryos and to a lesser extent in stomach cancer and liver.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, breast cancer; stomach cancer; embryonic defects; hepatic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
10 providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive and endocrine systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal  
15 fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for the diagnosis and/or treatment of a variety of disorders. Elevated expression  
20 of this gene product in stomach cancer indicates it is useful as a marker or therapeutic target for stomach cancer. Alternately, expression in breast tissue is influenced by the presence or absence of breast cancer tissue, and may thus also serve as a diagnostic marker for this cancer as well. Expression in the developing embryo may correlate with the normal development of human embryos, and expression in the liver is  
25 involved in the regulation of normal liver function and/or liver regeneration.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:55 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically  
30 excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general



formula of a-b, where a is any integer between 1 to 1079 of SEQ ID NO:55, b is an integer of 15 to 1093, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:55, and where b is greater than or equal to a + 14.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 46

This gene is expressed primarily in human hypothalamus derived from a patient with schizophrenia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, schizophrenia; neurological disorders; impaired nervous system function. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 169 as residues: Glu-34 to Trp-39. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain, particularly in the hypothalamus, indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal

cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated  
5 expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity,  
10 to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly  
15 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:56 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
20 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 618 of SEQ ID NO:56, b is an integer of 15 to 632, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:56, and where b is greater than or equal to a + 14.

## 25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 47**

The translation product of this gene shares sequence homology with human lecithin-cholesterol acyltransferase (LCAT), which catalyses the transfer of fatty acid from the sn-2 position of lecithin to the free hydroxyl group of cholesterol. Preferred polypeptides of the invention comprise the following amino acid sequence: RLVYN  
30 KTSRATQFPDGVDRVPGFGKTFSLFLDPSKSSVGSYFHTMVESLVGWGYT  
RGEDVRGAPYDWRRAPNENGPYFLALREMIEEMYQLYGGPVVLVAHSMGN  
MYTLYFLQRQPQAWKDKYIRAFVSLGAPWGGVAKTLRVLASGDNNRIPVIG

PLKIREQQRSAVSTSWLLPYNYSPEKVFVQTPTINYTLRDYRKFFQDIGFE  
DGWLMRQDTEGLVEATMPPGVQLHCLYGTGVPTPDSFYYESFPDRDPKICFG  
DGDGTVNLKSALQCQAWQSRQEHQVLLQELPGSEHIEMLANATTLAYLKRV  
LLGP (SEQ ID NO: 318). Polynucleotides encoding such polypeptides are also  
5 provided.

This gene is expressed primarily in osteoblasts & dendritic cells and to a lesser extent in muscle and other hematopoietic cell lineages.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
10 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hematopoietic disorders; immune dysfunction; osteoporosis; osteopetrosis; muscle degeneration. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential  
15 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,  
20 the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 170 as residues: Cys-65 to Ser-71. Polynucleotides encoding said polypeptides are also provided.

25 The tissue distribution and homology to lecithin-cholesterol acyltransferase (LCAT) indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of a variety of disorders. For example, arteriosclerosis is a pathological condition of mammals characterised by the accumulation of cholesterol in the arteries, which leads to heart disease, strokes, heart  
30 attacks and peripheral vascular disease. The enzyme could be used in a novel method of treating atherosclerosis, which involves increasing the level of LCAT activity, which then causes a decrease in the accumulation of cholesterol. The method and the

products can be used for the prophylaxis and treatment of atherosclerosis, and associated heart disease; myocardial infarction, stroke and peripheral vascular disease, as well as individuals suffering from Fish Eye Syndrome (caused by LCAT deficiency) or Classic LCAT Deficiency Syndrome. Alternately, elevated expression  
5 of this gene product in osteoblasts and hematopoietic cell lineages indicates that it may play additional roles in bone turnover, regulation of immune system function, and muscular function.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are  
10 related to SEQ ID NO:57 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
15 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2673 of SEQ ID NO:57, b is an integer of 15 to 2687, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:57, and where b is greater than or equal to a + 14.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 48**

20 When tested against HELA epithelial cell lines, supernatants removed from cells containing this gene activated the GAS assay. Thus, it is likely that this gene activates epithelial cells through the Jak-STAT signal transduction pathway. The gamma activating sequence (GAS) is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a  
25 large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in adult brain, infant brain, fibroblasts,  
30 embryonic and fetal tissue (e.g., spleen, liver), placenta and to a lesser extent in endocrine organs, cancerous colon and breast.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, dementia, epilepsy, schizophrenia, and developmental abnormalities.

5 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural system, endocrine system, and during development, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types  
10 (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15 The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly,  
20 the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive  
25 compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation,  
30 neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. In addition, the expression of this gene product in synovium (synovial sarcoma) would suggest a role in the detection and treatment of disorders and

conditions afflicting the skeletal system, in particular osteoporosis, bone cancer, connective tissue disorders (e.g. arthritis, trauma, tendonitis, chondromalacia and inflammation). The protein is also useful in the diagnosis or treatment of various autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, and dermatomyositis), dwarfism, spinal deformation, joint abnormalities, and chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid, etc.). The tissue distribution in endocrine tissues indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", and "Binding Activity" sections below, in Example 11, 17, 18, 19, 20 and 27, and elsewhere herein. Briefly, the protein can be used for the detection, treatment, and/or prevention of Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g. diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g. hyper-, hypothyroidism), parathyroid (e.g. hyper-, hypoparathyroidism), hypothalamus, and testes. Additionally, the expression within fetal tissue, cancerous colon and breast, and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the

polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of  
5 degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue  
10 markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly  
15 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:58 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
20 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 605 of SEQ ID NO:58, b is an integer of 15 to 619, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:58, and where b is greater than or equal to a + 14.

## 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 49

Preferred polypeptides of the invention comprise the following amino acid sequence or a subfragment thereof: MNKEDKVWNDCKGVNKLTNLEEQYIILIFQ  
NGLDPPANMVFESIINEIGIKNNISNFFAKIPFEEANGRLVACTRTYEESIKGSC  
GQKENKIKTVSFESKIQLRSKQEFQFFDEEEETGENHTIFIGPVEKLIVYPPPPA  
30 KGGISVTNEDLHCLNEGEFLNDVIIDFYLVLEKLKKEDADRIHIFSSFFYK  
RLNQRERRNHETTNLISIQQKRHGRVKTWTRHVDIFEKDFIFVPLNEAAHWFL  
AVVCFPGLEKPKYEPNPHYHENA VIQKCSTVEDSCISSASSEMESCSQNSSAK

PVIKKMLNKKHCIAVIDSNPGQEESDPYKRNICSVKYSVKKINHTASENEEF  
NKGESTSQKS (SEQ ID NO: 319). Polynucleotides encoding such polypeptides are also provided.

5 This gene is expressed primarily in fetal tissue, stomach, brain, endometrial cells, and bone and to a lesser extent in prostate, retina, adipocytes, smooth muscle, and tumors of the endometrium, ovaries, and parathyroid.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
10 not limited to, disorders of the endocrine system, ulcers, stomach cancer, epilepsy, schizophrenia, dementia, bone growth, developmental disorders and resorption. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the  
15 digestive system and neural systems expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, endocrine system, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression  
20 level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions.  
25 Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia,  
30 trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS,



psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Expression of this gene product in stomach tissue indicates involvement in digestion, processing, and elimination of food, as well as a potential role for this gene as a diagnostic marker or causative agent in the development of stomach cancer, and cancer in general. The expression within embryonic, fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. The tissue distribution in parathyroid tumor indicates polynucleotides and polypeptides corresponding to this

gene are useful for the detection, treatment, and/or prevention of various endocrine disorders and cancers. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", and "Binding Activity" sections below, in Example 11, 17, 18, 19, 20 and 27, and elsewhere herein. Briefly, the protein can be used for the detection, treatment, and/or prevention of Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g. diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g. hyper-, hypothyroidism), parathyroid (e.g. hyper-, hypoparathyroidism), hypothalamus, and testes. The tissue distribution in testes indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:59 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1364 of SEQ ID NO:59, b is an integer of 15 to 1378, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:59, and where b is greater than or equal to a + 14.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 50**

The translation product of this gene shares good protein homology with Xenopus NaDC-2 gene and a rabbit renal sodium/dicarboxylate cotransporter. The translation product of this gene also shares good homology with a rat placental protein which is a sodium-coupled high affinity dicarboxylate transporter. Therefore, it is likely that the translated product encoded by this gene shares similar biological activity.

This gene is expressed primarily in the placenta and colon adenocarcinoma. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental abnormalities as well as failure to thrive anomalies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system and colon, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., amniotic, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 173 as residues: Lys-166 to Gly-181. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in human placenta and the shared homology of this translation product to a rat placental protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of disorders of the placenta. Specific expression within the placenta indicates that this gene product may play a role in the proper establishment and maintenance of placental function. Alternately, this gene product is produced by the placenta and then transported to the embryo, where it may play a crucial role in the development and/or

survival of the developing embryo or fetus. Expression of this gene product in a vascular-rich tissue such as the placenta also indicates that this gene product is produced more generally in endothelial cells or within the circulation. In such instances, it may play more generalized roles in vascular function, such as in angiogenesis. It may also be produced in the vasculature and have effects on other cells within the circulation, such as hematopoietic cells. It may serve to promote the proliferation, survival, activation, and/or differentiation of hematopoietic cells, as well as other cells throughout the body. The tissue distribution in colon tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of disorders involving the colon. Expression of this gene product in colon tissue indicates involvement in digestion, processing, and elimination of food, as well as a potential role for this gene as a diagnostic marker or causative agent in the development of colon cancer, and cancer in general.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:60 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1112 of SEQ ID NO:60, b is an integer of 15 to 1126, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:60, and where b is greater than or equal to a + 14.

## **25 FEATURES OF PROTEIN ENCODED BY GENE NO: 51**

This gene is expressed primarily in the spinal cord.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, paralysis, neurologic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of

the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken  
5 from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in spinal cord indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or  
10 prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease,  
15 Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including  
20 disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or  
25 survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

30 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:61 and may have been publicly available prior to conception of

the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general  
5 formula of a-b, where a is any integer between 1 to 2064 of SEQ ID NO:61, b is an integer of 15 to 2078, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:61, and where b is greater than or equal to a + 14.

### **FEATURES OF PROTEIN ENCODED BY GENE NO: 52**

10 This gene is expressed primarily in keratinocytes, brain, fetal tissues, pericardium, stomach, and cancerous tissues (e.g., stomach, adrenals, parathyroid, germ cell, colon, breast).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
15 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, skin disorders, neurodegenerative and developmental disorders, heart disease, and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above  
20 tissues or cells, particularly of the cardiovascular and gastrointestinal systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the  
25 standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions.  
30 Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of

Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive  
5 compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation,  
10 neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. The tissue distribution in keratinocytes indicates polynucleotides and polypeptides corresponding to this gene are useful for the treatment, diagnosis, and/or prevention of various skin disorders. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", "infectious disease", and  
15 "Regeneration" sections below, in Example 11, 19, and 20, and elsewhere herein. Briefly, the protein is useful in detecting, treating, and/or preventing congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. keratoses, Bowen's Disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's Disease, mycosis  
20 fungoides, and Kaposi's sarcoma), injuries and inflammation of the skin (i.e. wounds, rashes, prickly heat disorder, psoriasis, dermatitis), atherosclerosis, urticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. In addition, such disorders may  
25 predispose increased susceptibility to viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, athlete's foot, and ringworm). Moreover, the protein product of this gene may also be useful for the treatment or diagnosis of various connective tissue disorders (i.e., arthritis, trauma, tendonitis, chondromalacia and  
30 inflammation, etc.), autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, dermatomyositis, etc.), dwarfism, spinal deformation, joint abnormalities, and chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita,

familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). The expression within fetal tissue (e.g., spleen and liver) and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Additionally, the tissue distribution in the pericardium of the heart indicates that the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the



protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:62 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 748 of SEQ ID NO:62, b is an integer of 15 to 762, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:62, and where b is greater than or equal to a + 14.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 53**

This gene is expressed primarily in the brain and in cartilage and to a lesser extent in the retina, activated T-cells, pineal gland, the lungs, and in synovial sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurological diseases, such as epilepsy and dementia, osteoarthritis, retinopathies, hematopoietic diseases, emphysema, and lung cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neurologic system, cartilage and musculature, vision, the hematopoietic system, and the pulmonary system expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, immune, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 176 as residues: Arg-34 to Cys-44. Polynucleotides encoding said polypeptides are also provided.

5 The tissue distribution in brain indicates polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of  
10 Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS,  
15 psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or  
20 survival. The tissue distribution in T-cells indicates polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of a variety of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product  
25 indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

30 Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such

as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The expression of this gene product in synovium would suggest a role in the detection and treatment of disorders and conditions afflicting the skeletal system, in particular osteoporosis, bone cancer, connective tissue disorders (e.g. arthritis, trauma, tendonitis, chondromalacia and inflammation). The protein is also useful in the diagnosis or treatment of various autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, and dermatomyositis), dwarfism, spinal deformation, joint abnormalities, and chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid, etc.). Additionally, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein: Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have

applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:63 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1080 of SEQ ID NO:63, b is an integer of 15 to 1094, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:63, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 54

This gene is expressed primarily in umbilical vein endothelial cells induced by IL-4.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are

not limited to, angiogenesis, inflammatory disorders, hematopoietic disease.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the angiogenic and hematopoietic systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in endothelial cells indicates polynucleotides and polypeptides corresponding to this gene are useful in the detection, treatment, and/or prevention of vascular conditions, which include, but are not limited to, microvascular disease, vascular leak syndrome, aneurysm, stroke, atherosclerosis, arteriosclerosis, or embolism. For example, this gene product may represent a soluble factor produced by smooth muscle that regulates the innervation of organs or regulates the survival of neighboring neurons. Likewise, it is involved in controlling the digestive process, and such actions as peristalsis. Similarly, it is involved in controlling the vasculature in areas where smooth muscle surrounds the endothelium of blood vessels. Furthermore,

the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues. The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, and as nutritional supplements. It may also have a very wide range of biological activities.

Representative uses are described in the "Chemotaxis" and "Binding Activity" sections below, in Examples 11, 12, 13, 14, 15, 16, 18, 19, and 20, and elsewhere herein. Briefly, the protein may possess the following activities: cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human

immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of hematopoiesis (e.g. for treating anemia or as adjunct to chemotherapy); stimulation or growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility);  
5 chemotactic and chemokinetic activities (e.g. for treating infections, tumors); hemostatic or thrombolytic activity (e.g. for treating hemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative diseases; for regulation of metabolism, and behavior. Also contemplated is the use of the  
10 corresponding nucleic acid in gene therapy procedures.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:64 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically  
15 excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1347 of SEQ ID NO:64, b is an integer of 15 to 1361, where both a and b correspond to the positions of nucleotide  
20 residues shown in SEQ ID NO:64, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 55

This gene is expressed primarily in both normal and cancerous pancreas.

Therefore, polynucleotides and polypeptides of the invention are useful as  
25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diabetes, gastrointestinal disorders, and pancreatic cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For  
30 a number of disorders of the above tissues or cells, particularly of the digestive and blood systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded

tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5       The tissue distribution in pancreas indicates that the protein products of this gene are useful as a therapeutic and/or diagnostic agent for pancreatic disorders and disorders of the endocrine and exocrine system, including but not limited to diabetes, blood disorders, pancreatic cancer, gastrointestinal diseases, hormonal imbalance, autoimmune disorders, cystic fibrosis, pancreatitis, and gallstones. Furthermore, the  
10       protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

15       Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:65 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is  
20       cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 933 of SEQ ID NO:65, b is an integer of 15 to 947, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:65, and where b is greater than or equal to a + 14.

## 25       **FEATURES OF PROTEIN ENCODED BY GENE NO: 56**

      The translation product of this gene shares sequence homology with oxidoreductase. Preferred polypeptides of the invention comprise the following amino acid sequence: MSPLSAARAALRVYAVGA AVILAQLLRRCRGGFLEPVXPPRP  
30       DRVAIVTGGTDGIGYSTANIWRDLGMHVIIAGNNSKAKQVVSKEETLND  
KVEFLYCDLASMTSIRQFVQKFKMKKIPLHVLINNAGVMMVPQRKTRDGFE  
HFGLNYLGHFLLTNLLD TLKESGSPGHSARVVTVSSATHYVAELNMDDLQS

SACYSPHAAYAQSKLALVLFTYHLQRLAAEGSHVTANVVDPGVVNTDXYK  
HVFWATRLAKKLLGWLLFKTPDEGAWTSIYAAVTPELEGVGGRYLYNEKET  
KSLHVTYNQKLQQQLWSKSCMTGVLDVTL (SEQ ID NO: 320). The mature  
form of this protein begins at residue 32. Thus, polypeptides comprising residues 2-  
330 and 32-330 of the sequence shown above are also provided. Polynucleotides  
encoding such polypeptides are also provided.

A preferred polypeptide fragment of the invention comprises the following  
amino acid sequence: MSPLSAARAALRVYAVGAAVILAQLLRRCRGGFLEP  
VXPPRPDRVAIVTGGTDGIG YSTANIWRDLACMLS (SEQ ID NO: 321).

Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in breast cancer cells, osteoclastoma, wilm's  
tumor, thymus stromal cells, and T cell helper I.

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions which include, but are  
not limited to, cancer, e.g., breast cancer, osteoclastoma, and wilm's tumor. Similarly,  
polypeptides and antibodies directed to these polypeptides are useful in providing  
immunological probes for differential identification of the tissue(s) or cell type(s). For  
a number of disorders of the above tissues or cells, particularly of the immune system,  
expression of this gene at significantly higher or lower levels is routinely detected in  
certain tissues or cell types (e.g., reproductive, kidney, immune, hematopoietic, and  
cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, breast milk,  
urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an  
individual having such a disorder, relative to the standard gene expression level, i.e.,  
the expression level in healthy tissue or bodily fluid from an individual not having the  
disorder.

The tissue distribution in breast cancer tissue, combined with the homology to  
oxidoreductase indicates that polynucleotides and polypeptides corresponding to this  
gene are useful for diagnosis and treatment of cancer, particularly, breast cancer,  
osteoclastoma, and wilm's tumor. This protein may play a role in the regulation of  
cellular division, and may show utility in the diagnosis, treatment, and/or prevention  
of developmental diseases and disorders, including cancer, and other proliferative



conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

5           Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have  
10 applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue  
15 differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the  
20 protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

25           Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:66 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is  
30 cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1362 of SEQ ID NO:66, b is an

integer of 15 to 1376, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:66, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 57

5           This gene is expressed primarily in monocytes, T cell helper II and B cell lymphoma.

          Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
10   not limited to, immune and hematopoietic diseases and/or disorders, particularly B-cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or  
15   lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not  
20   having the disorder.

          Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 180 as residues: Asp-30 to Val-40. Polynucleotides encoding said polypeptides are also provided.

          The tissue distribution in monocytes, T cell helper, and B cell lymphoma cells  
25   indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of B cell lymphoma. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or  
30   activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or

other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:67 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2420 of SEQ ID NO:67, b is an integer of 15 to 2434, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:67, and where b is greater than or equal to a + 14.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 58**

This gene is expressed primarily in human lung cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, pulmonary diseases and/or disorders, particularly cancers of the lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., pulmonary, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, pulmonary lavage, pulmonary surfactant, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 181 as residues: Phe-39 to Asp-45. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in lung cancer tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune system disorders such as ARDS, cystic fibrosis, and cancer, particularly lung cancer. This protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of

potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, 5 detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in 10 proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, 15 antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:68 and may have been publicly available prior to conception of 20 the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1072 of SEQ ID NO:68, b is an 25 integer of 15 to 1086, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:68, and where b is greater than or equal to a + 14.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 59**

This gene is expressed primarily in larynx carcinoma and early stage human 30 lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental, gastrointestinal, and pulmonary diseases and/or disorders, particularly larynx carcinoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., developmental, gastrointestinal, pulmonary, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, amniotic fluid, pulmonary lavage, sputum, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 182 as residues: His-42 to Lys-49. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in larynx carcinoma and early stage human lung indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating immune system disorders such as cancer, particularly larynx carcinoma. This protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the

polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of  
5 degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue  
10 markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly  
15 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:69 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
20 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1248 of SEQ ID NO:69, b is an integer of 15 to 1262, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:69, and where b is greater than or equal to a + 14.

## 25 **FEATURES OF PROTEIN ENCODED BY GENE NO: 60**

Preferred polypeptides of the invention comprise the following amino acid sequence: MEVTTEDTSRTDVSEPATSGGAADGVTSIAPTAVASSTTAASITTA  
ASSMTVASSAPTTAASSTTVASIAPTTTASSMTAASSTPMTLALPAPTSTXTGR  
TPSTTATGHPSLSTALAQVPKSSALPRTATLATLATRAQTVATTANTSSPMST  
30 RPSPSKHMPSDTAASPVPPMXPQAQGPISQVSVDQPVVNTTXKSTXMPSNTT  
XEPLTQAVVDKTL LLVLLLGVTLFITVLVLFALQAYESYKKKDYTQVDYLI

NGMYADSEM (SEQ ID NO: 322). Polynucleotides encoding these polypeptides are also provided.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by  
5 the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: ARCPGLPGLRCRPRPRAGPQAPSYCPRATRPPG ACCARMRLLEWRVYLRLTCATKDGMARECPTTWLSPPAKPDFAQRHVK PTALQGGRWSRLGASP (SEQ ID NO: 323). Polynucleotides encoding these polypeptides are also provided.

10 This gene is expressed primarily in adipocytes, osteoblasts, cerebellum, hypothalamus and Hodgkin's lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
15 not limited to, metabolic, skeletal, neural, and immune diseases and/or disorders, particularly Hodgkin's lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at  
20 significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., metabolic, skeletal, neural, immune, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily  
25 fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 183 as residues: Pro-33 to Gln-40, Gly-51 to Arg-56. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in Hodgkin's lymphoma cells indicates that  
30 polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune system disorders such as cancer, particularly Hodgkin's lymphoma. The secreted protein can also be used to determine biological activity, to



raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, and as nutritional supplements. It may also have a very wide range of biological activities. Representative uses are described in the "Chemotaxis" and "Binding Activity" sections below, in Examples 11, 12, 13, 14, 15, 16, 18, 19, and 20, and elsewhere herein. Briefly, the protein may possess the following activities: cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of hematopoiesis (e.g. for treating anemia or as adjunct to chemotherapy); stimulation or growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumors); hemostatic or thrombolytic activity (e.g. for treating hemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative diseases; for regulation of metabolism, and behavior. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:70 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1628 of SEQ ID NO:70, b is an integer of 15 to 1642, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:70, and where b is greater than or equal to a + 14.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 61**

The translation product of this gene shares sequence homology with polypeptide in the cystatin family. Cystatin polypeptides are cysteine protease inhibitors. For an analysis of the composition of several members of the cystatin family see Gene (1987) 61(3):329-338, incorporated herein by reference. The cystatin activity of polypeptides encoded by this gene is measured by several assays known in the art including assays described in coowned, copending US Patent Application Serial No. 08/744,138, incorporated herein by reference. Preferred polypeptides of the invention comprise the following amino acid sequence: LPATVEFAVHTFNQQSKD  
10 YYAYRLGHILNSWKEQVESKTVFSMELLGRTRCGKFEDDIDNCHFQUESTEL  
NNTFTCFFTISTRPWMTQFSLLNKTC (SEQ ID NO: 324). Fragments of such polypeptides having cystatin activity (cysteine protease inhibitory activity are particularly preferred). Polynucleotides encoding such polypeptides are also provided.

In another embodiment, polypeptides comprising the amino acid sequence of  
15 the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: LLWARGLGRAKSAVPTVST MLGLPWKGGLS  
WALLLLLLGSQILLIYAWHFHEQRDCDEHNVMARYLPATVEFAVHTFNQQS  
KDYYAYRLGHILNSWKEQVESKTVFSMELLGRTRCGKFEDDIDNCHFQE  
20 STELNNTFTCFFTISTRPWMTQFSLLNK TCLEGFH (SEQ ID NO: 325).  
Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in testes and epididymus. For a review of a cystatin showing testes- specific expression see Mol. Endocrinol. (1992 Oct.) 6(10):1653-1664, incorporated herein by reference.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, They should therefore serve a protective function to regulate the activities of such endogenous proteinases, which otherwise may cause uncontrolled  
30 proteolysis and tissue damage. Cysteine proteinase activity can normally not be measured in body fluids, but can be detected extracellularly in conditions like endotoxin-induced sepsis, metastasizing cancer, and at local inflammatory processes

in rheumatoid arthritis , purulent bronchiectasis and periodontitis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, testicular, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, seminal fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 184 as residues: Phe-31 to Asp-38, Asn-59 to Tyr-65, Ser-76 to Glu-82, Thr-96 to Cys-108, Gln-111 to Asn-118. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in testes and epididymus, combined with the homology to cystatins indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications.

Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent

of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Cysteine proteinase inhibitors of the cystatin superfamily are ubiquitous in the body and are generally tight-binding inhibitors of papain-like cysteine proteinases, such as cathepsins B, H, L, S, and K. They should therefore serve a protective function to regulate the activities of such endogenous proteinases, which otherwise may cause uncontrolled proteolysis and tissue damage. Cysteine proteinase activity can normally not be measured in body fluids, but can be detected extracellularly in conditions like endotoxin-induced sepsis, metastasizing cancer, and at local inflammatory processes in rheumatoid arthritis, purulent bronchiectasis and periodontitis, which indicates that a tight cystatin regulation is a necessity in the normal state. A deficiency state in which the levels of the intracellular cystatin, cystatin B, are lowered due to mutations has recently been shown to segregate with a form of progressive myoclonus epilepsy, which points to additional specialized functions of cystatins. Moreover, results showing that chicken cystatin inhibits polio virus replication, human cystatin C inhibits corona- and herpes simplex virus

replication, and human cystatin A inhibits rhabdovirus-induced apoptosis in cell cultures indicates that cystatins play additional roles in the human defense system. The cystatins constitute a superfamily of evolutionarily related proteins, all composed of at least one 100-120 residue domain with conserved sequence motifs.

5       The previously well characterized single-domain human members of this superfamily could be grouped in two protein families. The Family 1 members, cystatins (or stefins) A and B, contain approximately 100 amino acid residues, lack disulfide bridges, and are not synthesized as preproteins with signal peptides. The Family 2 cystatins (cystatins C, D, S, SN, and SA) are secreted proteins of approx.  
10   120 amino acid residues (Mr 13,000-14,000) and have two characteristic intrachain disulfide bonds. Recently, we identified an additional human cystatin superfamily member by EST1 sequencing in epithelial cell derived cDNA libraries which we named cystatin E. The same cystatin was independently discovered by differential display experiments as a mRNA species down-regulated in breast tumor tissue, but  
15   present in the surrounding epithelium and reported under the name cystatin M. Cystatin E/M is an atypical, secreted low-Mr cystatin in that it is a glycoprotein and just shows 30-35% sequence identity in alignments with the human Family 2 cystatins, which shows that additional cystatin families are yet to be identified. The cystatin E/M gene has been localized to chromosome 2, whereas all human Family 2  
20   cystatin genes are clustered on the short arm of chromosome 20, which further stresses that cystatin E/M is just distantly related to the other secreted human low-Mr cystatins. It is believed therefore, that polypeptides encoded by this gene are useful in diagnosing and treating disease consistent with the aforementioned conditions in which cystatins are implicated.

25       Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:71 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is  
30   cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 907 of SEQ ID NO:71, b is an

integer of 15 to 921, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:71, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 62

- 5       The translation product of this gene shares sequence homology with Neutrophil Gelatinase-Associated Lipocalin which is thought to be important in immune regulation (See Genbank and Geneseq Accession Nos. emb|CAA58127.1, and US5627034, respectively; all references and information available through these accessions are hereby incorporated herein by reference; for example, Biochem. Biophys. Res. Commun. 202 (3), 1468-1475 (1994), and FEBS Lett. 314 (3), 386-388 (1992)).

- 15       In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: LEQKLELHRGGGRSRTSGSPGLQEFGTREERGE GEQRTGREFSGNGGRAVEAARMRLLCGLWLWLSLLKVLQAQTPTPLPLPP PMQSFQGNQFQGEWFVLGLAGNSFRPEHRALLNAFTATFELSDDGRFEVWN AMTRGQHCDTWSYVLIPAAQPGQFTVDHGVGRSWLLPPGTLTDLQFICLGRAQ GLSDDNIVFPDVTGXALDL XSLPWVAAPA (SEQ ID NO: 326). Polynucleotides encoding these polypeptides are also provided.

20       This gene is expressed primarily in epididymus and osteoclastoma.

- 25       Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, reproductive and skeletal diseases and/or disorders, particularly cancers such as osteoclastoma testicular cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, testicular, skeletal, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, seminal fluid, urine, synovial fluid and spinal fluid)

or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 185 as residues: Met-82 to Thr-90. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in epididymus and homology to neutrophil gelatinase-associated lipocalin indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of skin diseases and immune system disorders such as cancer, particularly osteoclastoma. The secreted protein can also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, and as nutritional supplements. It may also have a very wide range of biological activities. Representative uses are described in the "Chemotaxis" and "Binding Activity" sections below, in Examples 11, 12, 13, 14, 15, 16, 18, 19, and 20, and elsewhere herein. Briefly, the protein may possess the following activities: cytokine, cell proliferation/differentiation modulating activity or induction of other cytokines; immunostimulating/immunosuppressant activities (e.g. for treating human immunodeficiency virus infection, cancer, autoimmune diseases and allergy); regulation of hematopoiesis (e.g. for treating anemia or as adjunct to chemotherapy); stimulation or growth of bone, cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds, stimulation of follicle stimulating hormone (for control of fertility); chemotactic and chemokinetic activities (e.g. for treating infections, tumors); hemostatic or thrombolytic activity (e.g. for treating hemophilia, cardiac infarction etc.); anti-inflammatory activity (e.g. for treating septic shock, Crohn's Disease); as antimicrobials; for treating psoriasis or other hyperproliferative diseases; for regulation of metabolism, and behavior. Also contemplated is the use of the corresponding nucleic acid in gene therapy procedures. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

related to SEQ ID NO:72 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
5 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 892 of SEQ ID NO:72, b is an integer of 15 to 906, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:72, and where b is greater than or equal to a + 14.

## 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 63

The translation product of this gene was shown to have homology to colipase which plays an essential role in the intestinal fat digestion by anchoring lipase on lipid/water interfaces in the presence of bile salts (See Genbank Accession No. gb|AAA03513.1; all references and information available through this accession are  
15 hereby incorporated by reference herein).

This gene is expressed primarily in epididymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
20 not limited to, reproductive diseases and/or disorders, particularly epididymus-related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or  
25 lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, metabolic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, seminal fluid, bile, chyme, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from  
30 an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 186 as residues: Ile-40 to Cys-49, Arg-52 to Cys-57,



Ser-94 to Trp-99, Gly-105 to Gly-111. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in epididymus indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune system diseases and disorders of the epididymus. Polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions concerning proper testicular function (e.g. endocrine function, sperm maturation), as well as cancer. Therefore, this gene product is useful in the treatment of male infertility and/or impotence. This gene product is also useful in assays designed to identify binding agents, as such agents (antagonists) are useful as male contraceptive agents. Similarly, the protein is believed to be useful in the treatment and/or diagnosis of testicular cancer. The testes are also a site of active gene expression of transcripts that is expressed, particularly at low levels, in other tissues of the body. Therefore, this gene product is expressed in other specific tissues or organs where it may play related functional roles in other processes, such as hematopoiesis, inflammation, bone formation, and kidney function, to name a few possible target indications. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:73 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 666 of SEQ ID NO:73, b is an integer of 15 to 680, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:73, and where b is greater than or equal to a + 14.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 64**

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: MCVCKERGREKEGGVTPTMTSNFPFCTLILGI  
5 AQAQACPGCPGDWPGLGSGVGEGLHHIRTCRTPIPCSPAPAAACLGS  
ARLPCVLRLLWPVPANLSSPFRLEALHCSFWSSPLLPAHLAFFGFRDLLTDFL  
LAACLLTFQKTPLELPMVAVHLLVATPCYQMLDNLPLPSAAAN WC (SEQ ID  
10 NO: 327). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in melanocytes and placenta and to a lesser extent in bone marrow and many cells of the immune system, including B-cells, dendritic cells, and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as  
15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, skin cancer and disorders of the reproductive and immune systems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell  
20 type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive and immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues and cell types (e.g., reproductive tissue, hematopoietic tissue, melanocytes and cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, amniotic  
25 fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in melanocytes indicates that polynucleotides and  
30 polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the skin, the reproductive system, and the immune system, particularly cancers. Representative uses are described in the "Biological Activity",

"Hyperproliferative Disorders", "infectious disease", and "Regeneration" sections below, in Example 11, 19, and 20, and elsewhere herein. Briefly, the protein is useful in detecting, treating, and/or preventing congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. keratoses, Bowen's Disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's Disease, mycosis fungoides, and Kaposi's sarcoma), injuries and inflammation of the skin (i.e. wounds, rashes, prickly heat disorder, psoriasis, dermatitis), atherosclerosis, urticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. In addition, such disorders may predispose increased susceptibility to viral and bacterial infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, athlete's foot, and ringworm). Moreover, the protein product of this gene may also be useful for the treatment or diagnosis of various connective tissue disorders (i.e., arthritis, trauma, tendonitis, chondromalacia and inflammation, etc.), autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, dermatomyositis, etc.), dwarfism, spinal deformation, joint abnormalities, and chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:74 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

formula of a-b, where a is any integer between 1 to 1619 of SEQ ID NO:74, b is an integer of 15 to 1633, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:74, and where b is greater than or equal to a + 14.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 65

Preferred polypeptides of the invention comprise the following amino acid sequence: YLWGRPRLRMRAGTSPSAPWGEEKREKLGHKLPVALQGYHPWIL  
LECTVFWARVVLACFSLYLIRGPNCINRQPEPTYQKACNLDCSSDFGQER  
APAWELLGPESEQRLREYTAQGLQSLASSHRWRQFKTEGKMRGGASPLPWLI  
10 CFW LCSYKGS DNSLKPVVP GPTLCPQSLVSPSVHPSTRSASLGRHRAEAA  
(SEQ ID NO: 328). Polynucleotides encoding these polypeptides are also provided.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the  
15 following amino acid sequence: MPGILAGIPVKDLCLSLQGFRLLLLCVCPGWL  
SGWMGGQKGSPRIVDIG (SEQ ID NO: 329). Polynucleotides encoding these polypeptides are also provided. This gene maps to chromosome 15, accordingly, polynucleotides of the invention is used in linkage analysis as a marker for chromosome 15.

20 This gene is expressed primarily in brain and breast and to a lesser extent in the liver, pancreas, and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
25 not limited to, disorders affecting the brain and CNS, the reproductive system, or the immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, the reproductive system, and the immune  
30 system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., brain and other tissue of the nervous system, mammary tissue, endocrine tissue, hepatic tissue, reproductive tissue, cells

and tissue of the immune system, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 188 as residues: Met-37 to Ser-43. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in brain cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the central nervous system, the reproductive system, and the immune system, including cancers. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

related to SEQ ID NO:75 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
5 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1008 of SEQ ID NO:75, b is an integer of 15 to 1022, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:75, and where b is greater than or equal to a + 14.

#### 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 66

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: AKGEERKEAFSLKMOVQLSSEPISEGLMYLYLGV  
15 FFHLIYPGALSITTLGKHSHPPFTAEQNSTVWMEHTLFHQSPVASHLVCFQSF  
AFSE (SEQ ID NO: 330). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in the brain and the immune system, in particular T-cells.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders affecting the brain, such as Alzheimer's or disorders affecting the immune system, such as AIDS. Similarly, polypeptides and antibodies directed to  
25 these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and CNS and the immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, cells and  
30 tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in brain cells and tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the brain and CNS or disorders affecting the immune system. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:76 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1170 of SEQ ID NO:76, b is an

integer of 15 to 1184, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:76, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 67

5       The translation product of this gene shares sequence homology with penaeidin-2 which is thought to be a members of a new family of antimicrobial peptides from the hemolymph of shrimps *Penaeus vannamei*. The molecules display antimicrobial activity against fungi and bacteria with a predominant activity against Gram-positive bacteria.

10       In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: GPAHPASPPLMTLSLQLAELVHFVCAFAQSQWT GVYPMPPLKPTPLCFA CVPCR (SEQ ID NO: 331). Polynucleotides  
15       encoding these polypeptides are also provided.

      This gene is expressed primarily in spleen.

      Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
20       not limited to, immune and hematopoietic diseases and/or disorders, particularly disorders affecting the spleen, including bacterial and fungal infections. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic  
25       and immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues and cell types (e.g., immune, hematopoietic, and cells and tissue of the immune system, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard  
30       gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.



The tissue distribution in spleen and homology to the penaeidin family of antibiotics indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the spleen, especially fungal and bacterial infections. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

related to SEQ ID NO:77 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
5 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 298 of SEQ ID NO:77, b is an integer of 15 to 312, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:77, and where b is greater than or equal to a + 14.

## 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 68

Contact of cells with supernatant expressing the product of this gene has been shown to increase the permeability of the plasma membrane of THP-1 cells to calcium. Thus it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface  
15 of the plasma membrane of both monocytes, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating immune and hematopoietic cells and tissue cell types. Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium and sodium, as well as alter pH and membrane  
20 potential. Alterations in small molecule concentration can be measured to identify supernatants which bind to receptors of a particular cell.

Moreover, when tested in TF-1 cell lines, the protein product of this gene has been shown to alter the steady-state messenger RNA levels of the following genes: c-fos, c-jun, egr-1, b561, bcl-2, CD40, cyclin D2, GADPH, ICER, MAD3, p21, STAT3,  
25 ID3, and STAT-1. When tested in U937 cell lines, the protein product of this gene has been shown to alter the steady-state messenger RNA levels of the following genes: egr2, MKP1, ATF3, B562, cyclin D, cyclin D2, GATA3, MAD3, p21, TGF, DHFR, and JAK3. Based upon these results, it is anticipated that polynucleotides and polypeptides corresponding to this gene are useful as agonists or antagonists of the  
30 above referenced genes. Such activity is useful in therapeutic and/or diagnostic applications as referenced and more specifically discussed elsewhere herein.

In specific embodiments, polypeptides of the invention comprise the sequence:MLLEVYGDSISVTVAIPL (SEQ ID NO:332), MHSPCQSKAADGLGKSETE (SEQ ID NO: 333), and/or MLKSLGLSTN (SEQ ID NO: 334). Polynucleotides encoding these polypeptides are also provided.

5 In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: AQRLAEECFYMLLEVYGDSISVTVAIPLMHSP  
10 CQSKAADGLGKSETEMLKSLGLSTNMSPFHLLGLKVFLTWALTLAQICLY  
VLSNSHTIPLSLYLPFPSKSRMPDTLHLLVHSLPLVHSQVLPVKDVTIEWPLC  
QRCLGSTCH Q (SEQ ID NO: 335). Polynucleotides encoding these polypeptides are also provided.

The polypeptide of this gene has been determined to have a transmembrane  
15 domain at about amino acid position 11 - 27 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing amino acids 28 to 143 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type Ia membrane proteins.

20 This gene is expressed primarily in neutrophils and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders and/or diseases affecting the immune system. Similarly,  
25 polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues and cell types (e.g., immune, hematopoietic, cells and tissue  
30 of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 191 as residues: Pro-97 to Asp-104. Polynucleotides  
5 encoding said polypeptides are also provided.

The tissue distribution in neutrophils and T-cells, combined with the detected calcium flux biological activity indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the immune system. Representative uses are described in the "Immune  
10 Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or  
15 other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such  
20 as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic  
25 lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and  
30 in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their

interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly  
5 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:78 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
10 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1356 of SEQ ID NO:78, b is an integer of 15 to 1370, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:78, and where b is greater than or equal to a + 14.

#### 15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 69**

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: WIPRAAGIRHEVQVSLFQMFCFSSIFCSH

20 EHTHLPGTFWLFLFLFLILPPSCPCFLPFS LAIETVRWPCWHHPTSFELCY  
PGTSIYYASRGGPXPNSEX (SEQ ID NO: 336). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in neutrophils.

Therefore, polynucleotides and polypeptides of the invention are useful as  
25 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, diseases and/or disorders affecting the immune system, and neutrophils in particular. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the  
30 tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues and cell types (e.g., blood cells,

and cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in neutrophils indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting the immune system and neutrophils in particular. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as,

antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:79 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 354 of SEQ ID NO:79, b is an integer of 15 to 368, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:79, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 70

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: XNXKSPLTIGNKSWSTAVAAALELVDPPGCRNSARDSPELVHLGKGRPRKLMTYLCSSISLLLLKVHSSGHQDIRKAKSKVPRLLIIQCPQQRE (SEQ ID NO: 337). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders affecting smooth muscle tissue, particularly vascular conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of smooth muscle tissue expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial

fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 193 as residues: Ser-18 to Val-31. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution primarily in smooth muscle indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of disorders affecting smooth muscle tissue. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:80 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1074 of SEQ ID NO:80, b is an integer of 15 to 1088, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:80, and where b is greater than or equal to a + 14.

### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 71

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by



the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: GPEENLSPSTPSQMPTIWVKLCLLQVCHGLFP LLKHWSQPMPLCVTLAPVSYWL (SEQ ID NO: 338). Polynucleotides encoding these polypeptides are also provided.

5           This gene is expressed primarily in fetal heart, smooth muscle, and frontal cortex.

          Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
10       not limited to, muscular, vascular, or neural diseases and/or disorders, particularly defects or injury to cardiac muscle. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, expression of this gene at  
15       significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., muscular, vascular, neural, developmental, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy  
20       tissue or bodily fluid from an individual not having the disorder.

          The tissue distribution in fetal heart indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating defects to the heart either due to injury or congenital defects. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and  
25       conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Alternatively, polynucleotides and polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory  
30       conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection,

treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:81 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1848 of SEQ ID NO:81, b is an integer of 15 to 1862, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:81, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 72

The translation product of this gene shares sequence homology with adipose complement related protein which is thought to be important in regulating energy metabolism, insulin levels and fat stores. Moreover, the protein product of this gene has also been shown to have homology to the complement subcomponent C1q A-chain precursor and HP-25 protein (See Genbank and Geneseq Accession Nos. emb|CAA41664.1, dbj|BAA02352.1, and W98013; all references and information

available through this accession are hereby incorporated by reference herein). Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with complement proteins.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: PRVRKEPEAMQWLRVRESPGEATGHRVTMG  
TAALGPVWAALLLFLLMCEIPMVELTFDRAVASDCQRCCDSEDPLDPAHVSS  
ASSSGRPHALPEIRPYINITILKGDKGDPGPMGLPGYMGREGPGGEPGPGQSK  
GDKGEMGSPGAPCQKRFFAFSVGRKTALHSGEDFQTLLFERVFNLDGC  
FDMATGQFAAPLRGIYFFSLNVHSWNYKETVYVHIMHNQKEAVILYAQPS  
ERSIMQSQSVMLDLAYGDRVWVRLFKRQRENAIYSNDFDTYITFSGHLIKA  
EDD (SEQ ID NO: 339). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in placenta and, fetal kidney, and umbilical vein and to a lesser extent in fetal heart, fetal liver/spleen, microvascular endothelial cells and cancers of the lung and pharynx.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, vascular, renal, and reproductive diseases and/or disorders, particularly cancers of the lung and pharynx. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary and immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., vascular, renal, reproductive, immune, hematopoietic, pulmonary, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 195 as residues: Asp-36 to Asp-48, Ser-57 to His-62, Lys-77 to Gly-84, Met-92 to Gly-114, Gln-203 to Ile-209, Lys-231 to Tyr-239.

Polynucleotides encoding said polypeptides are also provided.

5       The tissue distribution in pharynx or lung, combined with the homology to adipose complement related proteins indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating cancers of the pharynx or lung by modifying the metabolic balance in such tissues. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular  
10 disorders and conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product  
15 may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional  
20 supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:82 and may have been publicly available prior to conception of  
25 the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1604 of SEQ ID NO:82, b is an  
30 integer of 15 to 1618, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:82, and where b is greater than or equal to a + 14.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 73**

The translation product of this gene shares sequence homology with a hypothetical 54.7 kD protein (F37A4.1) from *Caenorhabditis elegans* (SwissProt locus YPT1\_CAEEL, accession P41879). The protein product of this gene also has  
5 homology to the human NG26 which is thought to contain a human major histocompatibility complex class III and is involved in T-cell maturation (See Genbank Accession No. gb|AAD18079.1| (AF129756); all references and information available through this accession are hereby incorporated by reference herein; for example, J. Neurochem. 69 (6), 2516-2528 (1997)). Based on the sequence similarity,  
10 the translation product of this gene is expected to share at least some biological activities with nitric oxide synthase proteins.

Preferred polypeptides of the invention comprise the following amino acid sequence: MLYPGSVYLLQKALMPVLLQGQARLVEECNGRRAKLLACDGNE  
IDTMFVDRRGTAEPQGQKLVICCEGNAGFYEVGCVSTPLEAGYSVLGWNHP  
15 GFAGSTGVFPQNEANAMDVVVQFAIHLRGFQPQDIIYAWSIGGFTATWAA  
MSYPDVSAMILDASFDDL VPLALKVMPDSWRGLVTRTVRQHLNLNNAEQLC  
RYQGPVLLIRRTKDEIITTTVPEDIMSNRGNL LLLKLLQHRYPRVMAEEGLRV  
VRQWLEASSQLEEASIYSRWEVEEDWCLSVLRSYQAEHGPDPFWSVGEDMS  
ADGRRQLALFLARKHLHNFEATHCTPLPAQNFQMPWHL (SEQ ID NO: 340).

20 Polynucleotides encoding such polypeptides are also provided.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: VCPKWCRFLTMLGHCCYFWQVWPASEALAA  
25 GPTPSTGSSSPSWKQHIGTSLQKTRGSLPTTTLTSGAGQSTSTGKNPAAGR  
SLEGALPAGVWPCFAQSPCTGGQQTTP SSTGLRSCLVRSPATWW RTP (SEQ ID NO: 341). Polynucleotides encoding these polypeptides are also provided.

Preferred polypeptides of the invention comprise the following amino acid sequence: WIPRAAGIRHEIYREXDSEAPASVPETPTAVTAPHSSSWDTYYQ  
30 PRALEKHADSILALASVFWSISYSSPFAFFYL YRKGYLSLSKVVPFSHYAG  
TLLLLLAGVACXRGIGRW TNPQYRQFITILEATHRNQSSSENKRQLANYNFD  
FRSWPVDFHWEPPSSRKESRGGPSRRGVALLRPEPLHRGTADTLLNRVKKL

PCQITSYLVAHTLGRRMLYPGSVYLLQKALMPVLLQGQARLVEECNGRRAK  
LLACDGNIDTMFVDRRGTAEPQGQKLVICCEGNAGFYEVGCVSTPLEAGYS  
VLGWNHPGFAGSTGVFPQNEANAMDVVVQFAIHRLGFQPQDIIYAWSI  
GGFTATWAAMSYPDVSAMILDASFDDLVPALKVMPDSWRGLVTRTVRQ  
5 HLNLNNAEQLCRYQGPVLLIRRTKDEIITTTVPEDIMSNRGNLLLLKLLQHRY  
PRVMAEEGLRVVRQWLEASSQLEEASIYSRWEVEEDWCLSVLRSYQAEHGP  
DFPWSVGEDMSADGRRQLALFLARKHLHNFEATHCT PLPAQNFQMPWHL  
(SEQ ID NO: 342). Polynucleotides encoding these polypeptides are also provided. A  
preferred polypeptide variant of the invention comprises the following amino acid  
10 sequence: HERAXGPSRGHGELLSCVLGPRLYKIYRERDSERAPASVPETPTA  
VTAPHSSSWDTYYQP RALEKHADSILALASVFWSISYYSSPFAFFYLRYKGY  
LSLSKVVPFSHYAGTLLLLLAGV ACSEALAAGPTPSTGSSSPSWKQHIGTSLQ  
KTRGSLPTTTLTSGAGQSTSTGKNPAAGRSLEGALPAGVWPCFAQSPCTGG  
QQTPSSTGL RSCLVRSPATWW RTP (SEQ ID NO: 343). Polynucleotides  
15 encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome  
6. Accordingly, polynucleotides related to this invention are useful as a marker in  
linkage analysis for chromosome 6.

This gene is expressed primarily in cerebellum, pituitary, fetal liver, and  
20 primary dendritic cells and to a lesser extent in in a wide range of tissues and  
developmental stages (i.e. fetal and adult tissue, etc.).

Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions which include, but are  
25 not limited to, neural, developmental, and immune diseases and/or disorders,  
particularly those involving self recognition and T- and B-cell maturation, and cancer.  
Similarly, polypeptides and antibodies directed to these polypeptides are useful in  
providing immunological probes for differential identification of the tissue(s) or cell  
type(s). For a number of disorders of the above tissues or cells, particularly of the  
30 neural or hormonal system, expression of this gene at significantly higher or lower  
levels is routinely detected in certain tissues or cell types (e.g., neural, developmental,  
immune, hepatic, and cancerous and wounded tissues) or bodily fluids (e.g., serum,

plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 196 as residues: Thr-23 to Lys-34, Leu-41 to Ser-47, Ala-57 to Ala-68, Pro-89 to Gly-101, Pro-110 to Pro-117. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in developmental and immune cells, combined with the  
10 homology to the human major histocompatibility complex class III region, indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of cancer and other proliferative disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the  
15 expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

---

20 Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory  
25 bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that  
30 influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and

in the differentiation and/or proliferation of various cell types. Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to  
5 isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly  
10 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:83 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
15 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2020 of SEQ ID NO:83, b is an integer of 15 to 2034, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:83, and where b is greater than or equal to a + 14.

## 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 74

The translation product of this gene shares sequence homology with the br-1 protein from the snail nervous system (EMBL HPBR1GENE) which codes for nitric oxide synthetase and which is thought to be important in mediating a variety of cellular responses, including vasodilation. Preferred polypeptides of the invention  
25 comprise the following amino acid sequence: MFKRHQRLKKDSTQAEEDLSEQ EQNQLNVLKKHGYVVGRTFLYSEEQKDNIPFEFDADSLAFDMENDPVM GTHKSTKQVELTAQDVKDAHWFYDTPGITKENCILNLLTEKEVNIVLPTQSIV PRTFVLKPGMVLFLGAIGRIDFLQGNQSAWFTVVASNILPVHITSLDRADALY QKHAGHTLLQIPMGGKERMAGFPPLVAEDIMLKEGLGASEAVADIKFSSAG  
30 WVSVTPNFKDRLHLRGYTPEGTVLTVRPPLLPIYIVNIKGQRIKKSVAAYKTKKP PSLMYNVRKKKGKINV (SEQ ID NO: 344). Polynucleotides encoding such polypeptides are also provided.



A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MLPARLPFRLLSLFLRGSAPTAARHGLREPLLERRCAA ASSFQHSSSLGRELPYDPVDTEGFGEGGDMQERFLFPEYILDPEPQPTREKQL QELQQQQEEEEERQRQQRREERRQQNLRARSREHPVVGHPPDPALPPSGVNCSS 5 GCGAXLHCQDAGVPGYLPREKFLRTAEADGGLARTVCQRCWLLSHHRRALR LQVSREQYLELVSAALRXPGPSLVLYMVDLLDLPDALLPDLPALVGPKQLIV LGNKVDLLPQDAPGYRQRLRERLWEDCARAGLLLAPGTKGHSAPSRTSHR TGRIRIRRTGPAQWSGTCG (SEQ ID NO: 345). Polynucleotides encoding these polypeptides are also provided.

10 When tested against U937 cell lines, supernatants removed from cells containing this gene activated the GAS (gamma activating sequence) promoter element. Thus, it is likely that this gene activates myeloid cells through the JAK-STAT signal transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is 15 a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

This gene is expressed primarily in early stage human brain, smooth muscle, 20 and endometrial tumor and to a lesser extent in a variety of tissues representing many organs and developmental states.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are 25 not limited to, cardiovascular, vascular, and neural diseases and/or disorders, particularly congestive heart disease and neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory and 30 neural systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cardiovascular, vascular, neural, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,

urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 5 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 197 as residues: Phe-42 to Leu-48, Pro-53 to Asp-58, Pro-81 to Glu-123, Asp-256 to Trp-269, Gly-282 to Ser-306, Arg-333 to Gly-339, Arg-403 to Gln-425, Ser-446 to Asn-452, His-475 to Gln-480, Gly-592 to Met-597, Pro-635 to His-642, Lys-667 to Lys-672, Lys-678 to Ser-684. Polynucleotides  
10 encoding said polypeptides are also provided.

- The tissue distribution in smooth muscle and vascular tissues, combined with the homology to nitric oxide synthetase indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of congestive heart failure and neurological degenerative disorders. polynucleotides and  
15 polypeptides corresponding to this gene are useful for the detection, treatment, and/or prevention of neurodegenerative disease states, behavioral disorders, or inflammatory conditions. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection,  
20 treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder,  
25 learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

- Potentially, this gene product is involved in synapse formation,  
30 neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to

microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis.

Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents  
5 that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are  
10 related to SEQ ID NO:84 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general  
15 formula of a-b, where a is any integer between 1 to 2226 of SEQ ID NO:84, b is an integer of 15 to 2240, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:84, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 75

20 The translation product of this gene shares sequence homology with the human KE04p, in addition to an unidentified C.elegans gene.

The polypeptide of this gene has been determined to have a transmembrane domain at about amino acid position 9 - 25 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing amino acids 1 to 8  
25 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type II membrane proteins.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by  
30 the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: PSFRRERVETGGGGPVTHGTEGPFLPLPGGTRM NMTQARVLVAADVGLVAVLLYASIHKIEEGHLAVYYRGGALLTSPSGPGYH

IMLPFITFRSVQTTLQTDEVKNVPCGTSGGVMYIDRIE VVNMLAPYAVFDIV  
RNYTADYDKTLIFNKHHELNQFCSAHTLQEVYIELFDQIDENLKQALQKDL  
NLMAPGLTIQAVRVTKPKIPEAIRNFELMEAEKTKLLIAAQKQKVVEKEA  
ETERKKAVIEAEKIAQVAKIRFQQKVMEEKETEKRISEIEDAAFLAREKAKA  
5 DAEYYAAHKYATSNKHKLTPEYLELKKYQAIASNSKIYFGSNIPNMFVDSSC  
ALKYSD IRTGRESSLPSKEALEPSGENVIQNKESTG (SEQ ID NO: 346).

Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome  
10. Accordingly, polynucleotides related to this invention are useful as a marker in  
10 linkage analysis for chromosome 10.

This gene is expressed primarily in fetal tissue, including 8 week whole  
embryo, fetal liver spleen, nine week old early stage human, fetal heart, fetal liver,  
fetal lung, and placenta and to a lesser extent in a variety of cancers, and other normal  
tissues.

15 Therefore, polynucleotides and polypeptides of the invention are useful as  
reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions which include, but are  
not limited to, cancer and diseases of fetal development. Similarly, polypeptides and  
antibodies directed to these polypeptides are useful in providing immunological  
20 probes for differential identification of the tissue(s) or cell type(s). For a number of  
disorders of the above tissues or cells, particularly of the fetal tissues, especially the  
liver, expression of this gene at significantly higher or lower levels is routinely  
detected in certain tissues or cell types (e.g., developmental, hepatic, immune,  
hematopoietic, pulmonary, cardiovascular, and cancerous and wounded tissues) or  
25 bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another  
tissue or cell sample taken from an individual having such a disorder, relative to the  
standard gene expression level, i.e., the expression level in healthy tissue or bodily  
fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic  
30 epitopes shown in SEQ ID NO: 198 as residues: Leu-68 to Lys-74, Tyr-109 to Lys-  
115, Gln-200 to Val-205, Lys-207 to Lys-214, Glu-237 to Ile-244, Ala-271 to Thr-

279, Ser-317 to Ser-329, Gln-342 to Gly-348. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution of this gene (primarily fetal tissue and cancerous tissue, both of which are undergoing rapid growth) indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of cancer and disorders of fetal development. Moreover, the expression within fetal tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as,

antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:85 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1474 of SEQ ID NO:85, b is an integer of 15 to 1488, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:85, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 76

When tested against U937 and Jurkat cell lines, supernatants removed from cells containing this gene activated the GAS (gamma activating sequence) promoter element. Thus, it is likely that this gene activates myeloid and T-cells, and to a lesser extent in other immune cells and tissue cell types, through the JAK-STAT signal transduction pathway. GAS is a promoter element found upstream of many genes which are involved in the Jak-STAT pathway. The Jak-STAT pathway is a large, signal transduction pathway involved in the differentiation and proliferation of cells. Therefore, activation of the Jak-STAT pathway, reflected by the binding of the GAS element, can be used to indicate proteins involved in the proliferation and differentiation of cells.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: WSTGNASWEKKDNFILSADFEMMGLGNRR  
SMKSPPLVLAALVACIIVLGFNYWIASSRSVDLQTRIMELEGRVRRRAAERG  
AVELKKNEFQGELEKQREQLDKIQSSHNFQLESVNKLYQDEKAVLVNNITTG  
ERLIRVLQDQLKTLQRNYGRLQQDVLQFQKNQTNLERKFSYDLSQCINQMKE  
VKEQCEERIEEVTKKGNEAVASRDLSNNNDQRQQLQALSEPQPRLQAAGL

PHTEVPQGKGNVLGNSKSQTPAPSSSEVVLD SKRQVEKEETNEIQVVNEE  
PQRDRLPQEPGREQVVEDRPVGGRGFGGAGELGQTPQVQAALXVSQENPE  
MEGPERDQLVIPDGQEEEQEAAGEGRNQQLRGEDDYNMDENEAESETDKQ  
AALAGNDRNIDVFNVE DQKRDTINLLDQREKRNHTL (SEQ ID NO: 347).

5 Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 9. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 9.

This gene is expressed primarily in human endometrial tumor and other  
10 tumors and to a lesser extent in a variety of other healthy adult and fetal tissues

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental diseases and/or disorders, particularly cancer and other  
15 proliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endometrial tissue, cervix and uterus, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues  
20 or cell types (e.g., developmental, reproductive, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 199 as residues: Asn-6 to Lys-12, Leu-65 to Phe-70, Glu-73 to His-88, Gln-123 to Gln-135, Gln-142 to Leu-156, Arg-173 to Gly-181, Asp-189 to Gln-199, Ser-204 to Arg-209, Glu-219 to Gly-225, Gly-229 to Pro-238, Ser-246 to Asn-256, Glu-263 to Arg-276. Polynucleotides encoding said polypeptides  
30 are also provided.

The tissue distribution in endometrial tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of

endometrial, cervical and uterine cancer. Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement.

Protein, as well as, antibodies directed against the protein may show

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:86 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 3160 of SEQ ID NO:86, b is an integer of 15 to 3174, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:86, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 77

The translation product of this gene shares sequence homology with protein disulfide isomerase from *Acanthamoeba castellanii* (See Genbank Locus ACADISPROA accession L28174, genpep locus 456013) which is thought to be important in converting proteins into their native conformations. The protein product of this gene was also shown to have homology to a phospholipase C homologue



derived from a mast cell cDNA library (See Geneseq Accession No. R99411). All references and information available through these accessions are hereby incorporated by reference herein - for example, Gene 150 (1), 175-179 (1994).

Included in this invention as preferred domains are endoplasmic reticulum  
 5 targeting sequence domain and the thioredoxin family active site domain, which were identified using the ProSite analysis tool (Swiss Institute of Bioinformatics). Proteins that permanently reside in the lumen of the endoplasmic reticulum (ER) seem to be distinguished from newly synthesized secretory proteins by the presence of the C-terminal sequence Lys-Asp-Glu-Leu (KDEL) [1,2]. While KDEL is the preferred  
 10 signal in many species, variants of that signal are used by different species. This situation is described in the following table.

Signal	Species-----
15 KDEL	Vertebrates, Drosophila, Caenorhabditis elegans, plants
HDEL	Saccharomyces cerevisiae, Kluyveromyces lactis, plants
DDEL	Kluyveromyces lactis
ADEL	Schizosaccharomyces pombe (fission yeast)
20 SDEL	Plasmodium falciparum

The signal is usually very strictly conserved in major ER proteins but some minor ER proteins have divergent sequences (probably because efficient retention of these proteins is not crucial to the cell). Proteins bearing the KDEL-type signal are not simply held in the ER, but are selectively retrieved from a post-ER compartment by a  
 25 receptor and returned to their normal location. The consensus pattern is as follows: [KRHQSA]-[DENQ]-E-L>. Thioredoxins are small proteins of approximately one hundred amino- acid residues which participate in various redox reactions via the reversible oxidation of an active center disulfide bond. They exist in either a reduced form or an oxidized form where the two cysteine residues are linked in an  
 30 intramolecular disulfide bond. Thioredoxin is present in prokaryotes and eukaryotes and the sequence around the redox-active disulfide bond is well conserved. Bacteriophage T4 also encodes for a thioredoxin but its primary structure is not homologous to bacterial, plant and vertebrate thioredoxins. A number of eukaryotic

proteins contain domains evolutionary related to thioredoxin, all of them seem to be protein disulphide isomerases (PDI). PDI (EC 5.3.4.1) is an endoplasmic reticulum enzyme that catalyzes the rearrangement of disulfide bonds in various proteins. The various forms of PDI which are currently known are: - PDI major isozyme; a  
5 multifunctional protein that also function as the beta subunit of prolyl 4-hydroxylase (EC 1.14.11.2), as a component of oligosaccharyl transferase (EC 2.4.1.119), as thyroxine deiodinase (EC 3.8. 1.4), as glutathione-insulin transhydrogenase (EC 1.8.4.2) and as a thyroid hormone-binding protein - ERp60 (ER-60; 58 Kd microsomal protein). ERp60 was originally thought to be a phosphoinositide-specific  
10 phospholipase C isozyme and later to be a protease. - ERp72. - P5. All PDI contains two or three (ERp72) copies of the thioredoxin domain. The concensus pattern is as follows: [LIVMF]-[LIVMSTA]-x-[LIVMFYC]-[FYWSTHE]-x(2)-[FYWGTN]-C-[GATPLVE]-[PHYWSTA]-C-x(6)-[LIVMFYWT]. The two C's form the redox-active bond.

15 Preferred polypeptides of the invention comprise the following amino acid sequence: SLHRFVLSQAKDEL (SEQ ID NO: 348), FIKFFAPWCGHCKALAPTW (SEQ ID NO: 349), and/or FIKFYAPWCGHCKTLAPTW (SEQ ID NO: 350). Polynucleotides encoding these polypeptides are also provided.

Further preferred are polypeptides comprising the endoplasmic reticulum  
20 targeting sequence domain and thioredoxin family active site domain of the sequence referenced in Table for this gene, and at least 5, 10, 15, 20, 25, 30, 50, or 75 additional contiguous amino acid residues of this referenced sequence. The additional contiguous amino acid residues is N-terminal or C- terminal to the endoplasmic reticulum targeting sequence domain and thioredoxin family active site domain.  
25 Alternatively, the additional contiguous amino acid residues is both N-terminal and C-terminal to the endoplasmic reticulum targeting sequence domain and thioredoxin family active site domain, wherein the total N- and C-terminal contiguous amino acid residues equal the specified number. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with  
30 thioredoxin proteins. Such activities are known in the art, some of which are described elsewhere herein.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: RRGRGVPGPRGRRRLWSAACGHCQRLQPTWN  
5 DLGDKYNSMEXAKVYVAKVDCTAHS DVC SAQGV RGYPTLKLFPQGEAV  
KYQGPRDFQTLN WMLQTLNEEPVTPEPEVEPPSAPELKQGLYELSASN FELH  
VAQGDHFIKFFAPWCGHCKALAPTWEQLALGLEHSETVKIGKVDCTQHY  
ELCSGNQVRGYPTLLWFRDGKKVDQYKGKRDLESLREYVESQLQRTETGA  
TETVTPSEAPVLAAEPEADKGTVLALTENNFD D TIAEGITFIKFYAPWCGHC  
10 KTLAPTWEELSKKEFPGLAGVKIAEVDCTAERNICKYSV RGYPTLL LFRGGK  
KVSEHSGGRDLDS LHRFVLSQAKDEL (SEQ ID NO: 351). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in human chondrosarcoma and endothelial cells and to a lesser extent in a wide range of normal and diseased adult and fetal  
15 tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, chondrosarcoma and other cancers and proliferative disorders.

20 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., vascular, skeletal,  
25 developmental, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution in chondrosarcoma, combined with the homology to protein disulfide isomerase and phospholipase C indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of

chondrosarcoma and other cancers and proliferative disorders, and possibly as a reagent for in vitro production of proteins. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Moreover, the expression in endothelial cells indicates the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

related to SEQ ID NO:87 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
5 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2766 of SEQ ID NO:87, b is an integer of 15 to 2780, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:87, and where b is greater than or equal to a + 14.

## 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 78

This gene is expressed primarily in thyroid and thymus

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
15 not limited to, thyroid diseases including thyroid cancer and diseases of function including Grave's Disease, hyper- and hypo- thyroidism as well as Diseases of the thymus. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,  
20 particularly of the endocrine and immune systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., endocrine, immune, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene  
25 expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in thyroid cells and tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the thyroid and thymus. Representative uses are  
30 described in the "Biological Activity", "Hyperproliferative Disorders", and "Binding Activity" sections below, in Example 11, 17, 18, 19, 20 and 27, and elsewhere herein. Briefly, the protein can be used for the detection, treatment, and/or prevention of

Addison's Disease, Cushing's Syndrome, and disorders and/or cancers of the pancreas (e.g. diabetes mellitus), adrenal cortex, ovaries, pituitary (e.g., hyper-, hypopituitarism), thyroid (e.g. hyper-, hypothyroidism), parathyroid (e.g. hyper-, hypoparathyroidism), hypothalamus, and testes. Furthermore, the protein may also  
5 be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

10 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:88 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is  
15 cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1047 of SEQ ID NO:88, b is an integer of 15 to 1061, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:88, and where b is greater than or equal to a + 14.

20

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 79**

The translation product of this gene shares sequence homology with collagen which is thought to be important as a structural material in a variety of human tissues and products including hair, nails, muscle and bone.

25 A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MRPQGPAASPQRLRGLLLLLLLQLPAPSSASEIPKGKQK AHSGRGRWWTCIMECAYKGQQECLVETGALGPMAFRVHLGSQVGMDSKEK RGNV (SEQ ID NO: 352). Polynucleotides encoding these polypeptides are also provided.

30 This gene is expressed primarily in smooth muscle and to a lesser extent in 12 week old early stage human, epididymus, healing groin wound, synovial hypoxia,

stromal cells, ulcerative colitis, breast and 8 week old embryo, as well as a variety of other normal and diseased cell types from adult and fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer and other proliferative disorders as well as Diseases of smooth muscle. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscular system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., vascular, developmental, reproductive, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 202 as residues: Glu-32 to Glu-46, Pro-63 to Ala-71, Pro-81 to Lys-90, Ser-97 to Trp-111, Lys-130 to Ser-135, Leu-147 to Cys-154, Asp-179 to Asn-186, Ser-219 to Gly-229. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in smooth muscle and homology to collagen indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases of vascular diseases and/or disorders. Representative uses are described in the "Biological Activity", "Hyperproliferative Disorders", "infectious disease", and "Regeneration" sections below, in Example 11, 19, and 20, and elsewhere herein. Briefly, the protein is useful in detecting, treating, and/or preventing congenital disorders (i.e. nevi, moles, freckles, Mongolian spots, hemangiomas, port-wine syndrome), integumentary tumors (i.e. keratoses, Bowen's Disease, basal cell carcinoma, squamous cell carcinoma, malignant melanoma, Paget's Disease, mycosis fungoides, and Kaposi's sarcoma), injuries and inflammation of the skin (i.e. wounds, rashes, prickly heat disorder, psoriasis, dermatitis),

atherosclerosis, urticaria, eczema, photosensitivity, autoimmune disorders (i.e. lupus erythematosus, vitiligo, dermatomyositis, morphea, scleroderma, pemphigoid, and pemphigus), keloids, striae, erythema, petechiae, purpura, and xanthelasma. In addition, such disorders may predispose increased susceptibility to viral and bacterial  
5 infections of the skin (i.e. cold sores, warts, chickenpox, molluscum contagiosum, herpes zoster, boils, cellulitis, erysipelas, impetigo, tinea, athletes foot, and ringworm).

Moreover, the protein product of this gene may also be useful for the treatment or diagnosis of various connective tissue disorders (i.e., arthritis, trauma,  
10 tendonitis, chondromalacia and inflammation, etc.), autoimmune disorders (i.e., rheumatoid arthritis, lupus, scleroderma, dermatomyositis, etc.), dwarfism, spinal deformation, joint abnormalities, and chondrodysplasias (i.e. spondyloepiphyseal dysplasia congenita, familial osteoarthritis, Atelosteogenesis type II, metaphyseal chondrodysplasia type Schmid). Moreover, the protein is useful in the detection,  
15 treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or  
20 receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly  
25 available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:89 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
30 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1328 of SEQ ID NO:89, b is an



integer of 15 to 1342, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:89, and where b is greater than or equal to a + 14.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 80**

5           This gene is expressed primarily in immune cells and to a lesser extent in a wide variety of human tissues.

          Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
10       not limited to, T cell or B cell leukemia and various immunodeficiencies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in  
15       certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

---

20       Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 203 as residues: Gly-3 to Gln-9. Polynucleotides encoding said polypeptides are also provided.

          The tissue distribution in immune cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of  
25       immune system diseases such as immunodeficiencies and T cell and/or B cell leukemia. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell  
30       lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory  
5 bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's  
10 Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the  
15 protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

20 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:90 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is  
25 cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 756 of SEQ ID NO:90, b is an integer of 15 to 770, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:90, and where b is greater than or equal to a + 14.

30

**FEATURES OF PROTEIN ENCODED BY GENE NO: 81**

The translation product of this gene shares sequence homology with IgE receptor. See for example, Isolation and Characterization of cDNAs coding for the Beta Subunit of the High-affinity Receptor for Immunoglobulin E, Proc. Natl. Acad. Sci. U S A. (1988 Sep.) 85(17): 6483-6487. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with IgE receptor proteins. Such activities are known in the art, some of which are described elsewhere herein. IgE and its receptors are believed to have evolved as a mechanism to protect mammals against parasites. But other and intrinsically innocuous antigens can subvert this system to provoke an allergic response. For human populations in industrialized countries, allergy and asthma now represent a far greater threat than parasitic infection, and the main impetus for current studies of the IgE system is the hope of understanding and intervening in the aetiology of allergic diseases. The high-affinity receptor for immunoglobulin (Ig) E (Fc epsilon RI) on mast cells and basophils plays a key role in IgE-mediated allergies. Fc epsilon RI is composed of one alpha, one beta, and two gamma chains, which are all required for cell surface expression of Fc epsilon RI, but only the alpha chain is involved in the binding to IgE. Fc epsilon RI-IgE interaction is highly species specific, and rodent Fc epsilon RI does not bind human IgE. New homolog can be used to develop anti-allergic agents. FcR deliver signals when they are aggregated at the cell surface. The aggregation of FcR having immunoreceptor tyrosine-based activation motifs (ITAMs) activates sequentially src family tyrosine kinases and syk family tyrosine kinases that connect transduced signals to common activation pathways shared with other receptors. FcR with ITAMs elicit cell activation, endocytosis, and phagocytosis. The nature of responses depends primarily on the cell type. The aggregation of FcR without ITAM does not trigger cell activation. Most of these FcR internalize their ligands, which can be endocytosed, phagocytosed, or transcytosed. The fate of internalized receptor-ligand complexes depends on defined sequences in the intracytoplasmic domain of the receptors. The coaggregation of different FcR results in positive or negative cooperation. Some FcR without ITAM use FcR with ITAM as signal transduction subunits. The coaggregation of antigen receptors or of FcR having ITAMs with FcR having immunoreceptor tyrosine-based inhibition motifs (ITIMs)

negatively regulates cell activation. FcR therefore appear as the subunits of multichain receptors whose constitution is not predetermined and which deliver adaptative messages as a function of the environment.

The polypeptide of this gene has been determined to have four transmembrane domains at about amino acid position 51 - 67, 89 - 105, 119 - 135, and 190 - 206 of the amino acid sequence referenced in Table 1 for this gene. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type IIIa membrane proteins.

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: ETRVKTSLELLRTQLEPTGTVGNTIMTSQPVPN ETIIVLPSNVINFSQAEKPEPTNQGQDSLKKHLHAEIKVIGTIQILCGMMVLSL GIILASASFSPNFTQVTSTLLNSAYPFIGPFFFIISGSLSIATEKRLTKLLVHSSLV GSILSALSALVGFIILSVKQATLNPASLQCELDKNNIPTRSYVSYFYHDSLYTT DCYTAKASLAGXLSLMLICTLLEFCLAVLTAVLRWKQAYSDFPGSVLFLPH SYIGNSGMSSKMTHDCGYEELLTS (SEQ ID NO: 353). Polynucleotides encoding these polypeptides are also provided.

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MMVLSLGIILASASFSPNFTQVTSTLLNSAYPFIGPFFFI ISGSLSIATEKRLTKLLVHSSLVGSILSALSALVGFIILSVKQATLNPASLQCELDKNNIPTRSYVSYFYHDSLYTTDCYTAKASLAGXLSLMLICTLLEFCL AVLTAVLRWKQAYSDFPGSVLFLPHSYIGNSGMSSKMTHDCGYEELLTS (SEQ ID NO: 354). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in immune system tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune system diseases and/or disorders such as cancer. Similarly,

polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in  
5 certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 204 as residues: Gln-23 to Lys-39, Glu-150 to Thr-158. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in immune cells and tissues combined with the homology to IgE receptor indicates that polynucleotides and polypeptides  
15 corresponding to this gene are useful for diagnosis and treatment of immune system disorders. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell  
20 lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for  
25 immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity  
30 disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that

influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. The gene product may also be involved in lymphopoiesis, therefore, it can be used in immune disorders such as infection, inflammation, allergy, immunodeficiency etc. In addition, this gene product may have commercial utility in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:91 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1556 of SEQ ID NO:91, b is an integer of 15 to 1570, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:91, and where b is greater than or equal to a + 14.

## **25 FEATURES OF PROTEIN ENCODED BY GENE NO: 82**

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: GASCEGGGAAARAALGVHRSQKALLVFRRTL  
SNLLYMPLLRGLLWLQVLCAGPLHTEAVVLLVPSDDGRAFLLRSLHPEAH  
VPPAADRGASLQCVLHQAAPKSRPRSPAAGAALLHXPRRTGDEPCREFHGN  
GFPGPTQLTPGECGLPAPSSLLQHASAPVRTGSEGQVVGCPRARGETGEGLSL

AFLSSLMFTSRNGLVGC GASCEGGGAAARAALGVHRSQKALLVFRRTLNL  
LYMPLLRLGLLWLQVLCAGPLHTEAVVLLVPSDDGRAFLLSRLLHPEAHVPP  
AAD RGASLQCVLHQAAPKSRPRSPAAGAALLHXPRRTGDEPCREFHGNGFP  
GPTQLTPGECGLPAPSSLLQHASAPVRTGSEGQVVGCPRARGETGEGLSLA

5 FLSSLMFTSRNGLVGC (SEQ ID NO: 355). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 7. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 7.

10 This gene is expressed primarily in activated T cells, and to a lesser extent in a wide variety of human tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
15 not limited to, immune and hematopoietic diseases and/or disorders, particularly immunodeficiencies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at  
20 significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
25 individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 205 as residues: Pro-67 to Ser-73. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in activated T cells indicates that polynucleotides and  
30 polypeptides corresponding to this gene are useful for diagnosis and treatment of immunodeficiencies. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and

elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:92 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general



formula of a-b, where a is any integer between 1 to 2936 of SEQ ID NO:92, b is an integer of 15 to 2950, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:92, and where b is greater than or equal to a + 14.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 83

The translation product of this gene was shown to have homology to the human transmembrane protein (See Genbank Accession No. gblAAC51364.11 (AF000959); all references and information available through this accession are hereby incorporated by reference herein; for example, Genomics 42 (2), 245-251  
10 (1997)) which is thought to be implicated in velo-cardio-facial syndrome.

A preferred polypeptide fragment of the invention comprises the following amino acid sequence: MGSAALEILGLVLCLVGWGGILACGLPMWQVTAFLD  
HNIVTAQTTWKGLWMSCVVQSTGTCSAKCTTRCWL (SEQ ID NO: 356).

Polynucleotides encoding these polypeptides are also provided.

15 The gene encoding the disclosed cDNA is believed to reside on chromosome 22. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 22.

This gene is expressed primarily in dementia brain tissue, and to a lesser extent in a wide variety of human tissues.

---

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neural diseases and/or disorders, particularly dementia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing  
25 immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or  
30 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 206 as residues: Ser-201 to Tyr-217. Polynucleotides encoding said polypeptides are also provided.

5 The tissue distribution in dementia brain tissue, combined with the homology to the transmembrane protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of dementia, and potentially for velo-cardio-facial syndrome. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the 10 detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, 15 learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, 20 neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show 25 utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:93 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically 30 excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general

formula of a-b, where a is any integer between 1 to 1708 of SEQ ID NO:93, b is an integer of 15 to 1722, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:93, and where b is greater than or equal to a + 14.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 84

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: LKRAPPGPALAKGLLQPSSTFQALETNIGDQVR  
10 RHSTAVVIREMTSYILISFVLLIGVGCIEKDQSCPVFGGRKRLHLLFVGGQLRQ  
VRMLRGELSCACYRPHVQALQLGGCTCF (SEQ ID NO: 357). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in the adult pulmonary system.

Therefore, polynucleotides and polypeptides of the invention are useful as  
15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cystic fibrosis, bronchitis and any pulmonary disorders in general. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell  
20 type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., pulmonary, cardiovascular, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, pulmonary surfactant, pulmonary lavage/sputum, synovial fluid and spinal fluid) or  
25 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene only in the pulmonary system indicates that it plays a key role in the functioning of the pulmonary system. This would suggest  
30 that misregulation of the expression of this protein product in the adult could lead to lymphoma or sarcoma formation, particularly in the lung and the protein product could be used either in the treatment and/or detection of these disease states. The gene

or gene product may also be useful in the treatment and/or detection of pulmonary defects such as pulmonary edema and embolism, bronchitis and cystic fibrosis.

Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents  
5 that modulate their interactions, in addition to its use as a nutritional supplement.

Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are  
10 related to SEQ ID NO:94 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
15 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 621 of SEQ ID NO:94, b is an integer of 15 to 635, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:94, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 85

20 The translation product of this gene was found to be homologous to CAM proteins. Based on the sequence similarity, the translation product of this gene is expected to share at least some biological activities with CAM proteins. Such activities are known in the art, some of which are described elsewhere herein.

A preferred polypeptide variant of the invention comprises the following  
25 amino acid sequence: MLCPWRTANLGLLLILTIFLVAEAEGAAQPNNSLM  
LQTSKENHALASSSLCMDEKQITQNYSKVLAEVNTSWPVKMATNAVLC  
CPPIALRNLIITWEIILRGQPSCTKAYKKETNETKETNCTDERITWVSRPDQ  
NSDLQIRTVAITHDGYYRCIMVTPDGNFHRGYHLQVLVTPEVTLFQNRNRTA  
VCKAVAGKPAAHISWIPEGDCATKQEYWSNGTVTVKSTCHWEVHNVSTV  
30 NCHVSHLTGNKSLYIELLPVPGAKKSSKLYIPYIILTIILTIVGXIWLLKVNG  
CXKYKLNKPESTPVVEEDEMOPYAFYTEKNNPLXXTTNKVKASEALQSEV

DTDLHTL (SEQ ID NO:208). Polynucleotides encoding these polypeptides are also provided.

5 The polypeptide of this gene has been determined to have a transmembrane domain at about amino acid position 271 - 287 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing amino acids 288 to 348 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type Ia membrane proteins.

This gene is expressed primarily in dendritic cells.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immunodeficiency, tumor necrosis, infection, lymphomas, auto-immunities, cancer, metastasis, wound healing, inflammation, anemias  
15 (leukemia) and other hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell  
20 types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 208 as residues: Asp-53 to Tyr-61, Pro-105 to Ile-128, Arg-133 to Leu-140, Gln-182 to Ala-188, Pro-205 to Asn-218, Gly-259 to Ala-264, Asn-290 to Ser-302, Glu-307 to Tyr-314, Tyr-317 to Lys-332. Polynucleotides encoding said polypeptides are also provided.

30 The tissue distribution in dendritic cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of immune disorders including: leukemias, lymphomas, auto-immunities,

immunodeficiencies (e.g. AIDS), immuno-suppressive conditions (transplantation) and hematopoietic disorders. In addition this gene product is applicable in conditions of general microbial infection, inflammation or cancer. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in  
5 Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the  
10 treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory  
15 bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's  
20 Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the  
25 protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

30 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:95 and may have been publicly available prior to conception of

the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general  
5 formula of a-b, where a is any integer between 1 to 3784 of SEQ ID NO:95, b is an integer of 15 to 3798, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:95, and where b is greater than or equal to a + 14.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 86**

10 In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: VIKLICPAAFPVYFQDMARGCVCSLCASVCIFLS  
SLFPLLPSVHSVNIISCLLLSKCFEGLELMCEHL YQLSQLHVLHHIFS YLLCTP  
15 (SEQ ID NO: 358). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in embryonic tissue and to a lesser extent in in a variety of other tissues and cell types.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, developmental anomalies, fetal deficiencies, cancer and neoplastic states. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the  
25 developing fetus, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., developmental, differentiating, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, amniotic fluid, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression  
30 level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in embryonic tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of developmental anomalies, fetal deficiencies and pre-natal disorders, as well as abnormal cell proliferation and/or differentiation, neoplastic states and cancer.

5 Moreover, the expression within embryonic tissue and other cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders"

10 and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent

15 of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the

20 polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in

25 modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their

30 interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.



Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:96 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically  
5 excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2669 of SEQ ID NO:96, b is an integer of 15 to 2683, where both a and b correspond to the positions of nucleotide  
10 residues shown in SEQ ID NO:96, and where b is greater than or equal to a + 14.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 87**

The translation product of this gene shares sequence homology with inter-alpha-trypsin inhibitor which is thought to be important in inhibition of trypsin and  
15 other serine proteases (See Genbank Accession No. pirlS30350IS30350; all references and information available through this accession are hereby incorporated herein by reference; for example, Eur. J. Biochem. 179 (1), 147-154 (1989), J. Biol. Chem. 264 (27), 15975-15981 (1989), and J. Biol. Chem. 266 (2), 747-751 (1991)).

Contact of cells with supernatant expressing the product of this gene has been  
20 shown to increase the permeability of the plasma membrane of THP-1 cells to calcium. Thus it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of both monocytes, in addition to other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are  
25 not limited to, activating monocytes, and to a lesser extent, other immune and/or hematopoietic cells. Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium and sodium, as well as alter pH and membrane potential. Alterations in small molecule concentration can be measured to identify supernatants which bind to receptors of a particular cell.

30 In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the

following amino acid sequence: YXIPGSTHASGRQRGSGRGEDDSGPPPSTVINQ  
NETFANIIFKPTVVQQARIAQNGILGDFIIRYDVNREQSIGDIQVLNGYFVHYF  
APKDLPLPKNVVFLDSSASMVGTKLRQTKDALFTILHDLRPQDRFSIIGFS  
NRIKVWKDHLISVTPDSIRDGKVYIHHMSPTGGTDINGVLQRAIRLLNKYVAH  
5 SGIGDRSVSLIVFLTDG KPTVGETHTLKLNNNTREAARGQVCIFTIGIGNDVD  
FRLLEKLSLENCGLTRRVHEEEDAGSQLIGFYDEIRTPLLSDIRIDYPPSSVVQ  
ATKTLFPNYFNGSEIIIAGKLVDRKLDHLHVEVTASNSKKFIILKTDVPVRPQK  
AGKDVTGSPRPGDGEGDXNHIERLWSYLTTKELLSSWLQSDDEPEKERLRQ  
RAQALAVSYRFLTPFTSMKLRGPVPRMDGLEEAHGMSAAMGPEPVVQSVR  
10 GAGTQPGPLLKKPYQPRIKISKTSVDGDPHFVVDPLSRLTVCFNIDGQPGDIL  
RLVSDHRDSGVTVNGELIGAPAPPNGHKKQRTYLRTITILINKPERSYLEITPS  
RVILDGGDRLVLPCNQSVVVGSWGLEVSVSANANVTVTIQGSIAFVILIHLYK  
KPAFQRHHLGFYIANSEGLSSNCHGLLGQFLNQDARLTEDPAGPSQNLTHP  
LLLQVGEGPEAVLTVKGHQVPVWVKQRKIYN GEEQXDCWFARNMPPN

15 (SEQ ID NO: 359). Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in placenta and adipose tissue and to a lesser extent in several other organs and tissues including cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
20 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of developing organs and metabolic diseases, in addition to vascular diseases and conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential  
25 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the developing systems and metabolic systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., reproductive, vascular, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or  
30 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 210 as residues: Lys-5 to Lys-10, Asn-33 to Lys-39, Asp-48 to Lys-54, Pro-62 to Asp-67, Asn-116 to Arg-123, His-157 to Ala-162, Val-242 to Lys-249, Val-251 to Asp-264. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in placenta, combined with the homology to inter-alpha-trypsin inhibitor and the detected calcium flux biological activity indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders of developing and metabolic systems. This protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease,

vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Polynucleotides and polypeptides of the invention are also useful for the treatment, detection, and/or prevention of inflammation, tumor invasion and metastasis, wound healing, liver disease, disseminated intravascular coagulation, alzheimer's Disease, ophthalmic disease, apoptosis, tissue remodeling, intrauterine growth retardation, preeclampsia, angiogenesis, cell migration, fetal development, trophoblast implantation, ovulation, pemphigus and psoriasis, and antiviral therapy. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:97 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 2167 of SEQ ID NO:97, b is an integer of 15 to 2181, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:97, and where b is greater than or equal to a + 14.

## **25 FEATURES OF PROTEIN ENCODED BY GENE NO: 88**

The translation product of this gene was shown to have homology to the human colon carcinoma antigen NY-CO-7 (See Genbank and Geneseq Accession Nos. gb|AAC18038.1| (AF039689) and WO9904265; all references available through this accession are hereby incorporated herein by reference; for example, Int. J. Cancer 76 (5), 652-658 (1998)).

This gene is expressed primarily in breast and breast cancer and to a lesser extent in several other organs and tissues including cancers.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of reproductive organs and the gastrointestinal system, including cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., gastrointestinal, reproductive, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, breast milk, chyme, bile, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 211 as residues: Gly-22 to Gly-28, Leu-71 to Phe-77, Asn-101 to Val-108, Pro-122 to Ser-127, Arg-149 to Pro-154, Gly-191 to Phe-196, Pro-199 to Thr-211. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in breast and breast cancer tissue, combined with the homology to a colon cancer antigen indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of disorders of the reproductive systems and cancers. This protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of

potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, 5 detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in 10 proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, 15 antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:98 and may have been publicly available prior to conception of 20 the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1943 of SEQ ID NO:98, b is an 25 integer of 15 to 1957, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:98, and where b is greater than or equal to a + 14.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 89**

The translation product of this gene shares sequence homology with the amino 30 acid and protein sequence of a *Xenopus* transmembrane protein of unknown function. The very 5'-end of the contig is identical to the mRNA for the human LGN mosaic protein. Based on the sequence similarity, the translation product of this gene is

expected to share at least some biological activities with LGN mosaic proteins. Such activities are known in the art, some of which are described elsewhere herein.

Preferred polypeptides of the invention comprise the following amino acid sequence:

PRVRPPTKALAVTFTTFVTEPLKHIGKGTGEFIKALMKEIPALLHLPVLIIMAL  
5 AILSFCYGAGKSVHVLRHIGGPEREPPQALRPRDRRRQEEIDYRPDGGAGDAD  
FHYRGQMGPTEQGPYAKTYEGRREILRERDVDLRFQTGNKSPEVLRAFDVPD  
AEAREHPTVVPSHKSPVLDTKPKETGGILGEGTPKESSTESSQSAKPVSGQDTS  
GNTEGSPA AEKAQLKSEAAGSPDQGSTYSPARGVAGPRGQDPVSSPCG (SEQ  
ID NO:339). Polynucleotides encoding such polypeptides are also provided.

10 The gene encoding the disclosed cDNA is believed to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

This gene is expressed primarily in small intestine and adipocytes and to a lesser extent in various other normal and transformed cell types, mostly of endocrine origin.  
15

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, conditions of growth and metabolism. Similarly, polypeptides and  
20 antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive and endocrine systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., metabolic, gastrointestinal, and  
25 cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, bile, chyme, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 212 as residues: Pro-40 to Gly-68, Gly-79 to Arg-93, Phe-106 to Glu-114, Pro-122 to His-129, Thr-143 to Gly-149, Gly-155 to Ala-168,

Val-171 to Gly-182, Ala-195 to Pro-207, Pro-214 to Val-220. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in small intestine indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of disorders of growth and metabolism as well as endocrine abnormalities. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:99 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1098 of SEQ ID NO:99, b is an integer of 15 to 1112, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:99, and where b is greater than or equal to a + 14.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 90**

The translation product of this gene shares sequence homology with IgE receptor beta chain which is thought to be important in immune function.

This gene is expressed primarily in kidney medulla tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune and renal diseases and/or disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and renal systems,



expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, renal, urogenital, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such  
5 a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in kidney renal medulla tissue, combined with the homology to the IgE receptor beta chain indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of immune and  
10 renal disorders. The protein product of this gene could be used in the treatment and/or detection of kidney diseases including renal failure, nephritis, renal tubular acidosis, proteinuria, pyuria, edema, pyelonephritis, hydronephritis, nephrotic syndrome, crush syndrome, glomerulonephritis, hematuria, renal colic and kidney stones, in addition to  
15 Wilm's Tumor Disease, and congenital kidney abnormalities such as horseshoe kidney, polycystic kidney, and Falconi's syndrome. Alternatively, this gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene  
20 product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and  
25 tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits  
30 hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the

protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:100 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 873 of SEQ ID NO:100, b is an integer of 15 to 887, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:100, and where b is greater than or equal to a + 14.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 91**

The translation product of this gene shares sequence homology with Diff 40 gene product (See Genbank Accession No. gblAAC51134.1; all references and information available through this reference are hereby incorporated herein).

Preferred polypeptides of the invention comprise the following amino acid sequence: PRVRSIKVTELKGLANHVVVGSVSCETKDLFAALPQVVAVDIN  
DLGTIKLSLEVTWSPFDKDDQPSAASSVNKASTVTKRFSTYSQSPDTPS  
LREQAFYNMLRRQEELENGTAWSLSSSESSDDSSSPQLSGTARHSPAPRPLV  
QQPEPLPIQVAFRRPETPSSGPLDEEGAVAPVLANGHAPYSRTLHISEASVNA  
ALAEASVEAVGPKSLSWGSPPTHAPATHGKHPSVPVPPALDPGHSATSST  
LGTTGSVPTSTDPAPSAHLDSVHKSTDGSPSELPGPTHTTTGSTYSAITTTHS  
APSPLTHTTTGSTHKPIISTLTTTGPTLNIIGPVQTTTSPHTTMPSPSSHNSNPQ  
YVDFCSSVCDNIFVHYVIGIFFHTLYSSKTL (SEQ ID NO:360), and/or PRVRS  
IKVTELKGLANHVVVGSVSCETKDLFAALPQVVAVDINDLGTIKLSLEVTWSP  
FDKDDQPSAASSVNKASTVTKRFSTYSQSPDTPSLREQAFYNMLRRQEELE  
NGTAWSLSSSESSDDSSSPQLSGTARHSPAP RPLVQQPEPLPIQVAFRRPET

PSSGPLDEEGAVAPVLANGHAPYSRTLSEASVNAALAEASVEAVGPKSL  
SWGSPPTHAPATHGKHPSVPPALDPGHSATSSTLGTGTVPTSTD (SEQ ID  
NO: 361). Polynucleotides encoding these polypeptides are also provided.

Polypeptides of the invention do not consist of the primary amino acid sequence  
5 shown as Geneseq Accession No.W69430, which is hereby incorporated herein by  
reference.

This gene is expressed primarily in liver and to a lesser extent in gall bladder  
tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as  
10 reagents for differential identification of the tissue(s) or cell type(s) present in a  
biological sample and for diagnosis of diseases and conditions which include, but are  
not limited to, metabolic and endocrine diseases and/or disorders, particularly hepatic  
and gall bladder disorders. Similarly, polypeptides and antibodies directed to these  
polypeptides are useful in providing immunological probes for differential  
15 identification of the tissue(s) or cell type(s). For a number of disorders of the above  
tissues or cells, particularly of the metabolic and endocrine systems, expression of this  
gene at significantly higher or lower levels is routinely detected in certain tissues or  
cell types (e.g., hepatic, metabolic, gall bladder, gastrointestinal, and cancerous and  
wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, bile, synovial  
20 fluid and spinal fluid) or another tissue or cell sample taken from an individual having  
such a disorder, relative to the standard gene expression level, i.e., the expression  
level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic  
epitopes shown in SEQ ID NO: 214 as residues: Val-9 to Cys-14, Pro-42 to Thr-47,  
25 Thr-56 to Ala-64, Asp-88 to His-98, Cys-128 to Ser-136, Arg-153 to Trp-161.  
Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in liver and gall bladder, combined with the homology  
to the diff 40 gene product indicates that polynucleotides and polypeptides  
corresponding to this gene are useful for the study and treatment of endocrine and  
30 metabolic disorders. polynucleotides and polypeptides corresponding to this gene are  
useful for the detection and treatment of liver disorders and cancers. Representative  
uses are described in the "Hyperproliferative Disorders", "infectious disease", and

"Binding Activity" sections below, in Example 11, and 27, and elsewhere herein. Briefly, the protein can be used for the detection, treatment, and/or prevention of hepatoblastoma, jaundice, hepatitis, liver metabolic diseases and conditions that are attributable to the differentiation of hepatocyte progenitor cells. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:101 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1234 of SEQ ID NO:101, b is an integer of 15 to 1248, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:101, and where b is greater than or equal to a + 14.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 92

The polypeptide of this gene has been determined to have a transmembrane domain at about amino acid position 3 - 19 of the amino acid sequence referenced in Table 1 for this gene. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type II membrane proteins.

This gene is expressed primarily in fetal brain and to a lesser extent in pancreas tumor, melanocyte and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neural diseases and/or disorders, particularly neurodevelopmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are

useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., neural, developmental, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution in fetal brain tissue indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of developmental disorders of the central nervous system. Representative uses are described in the "Regeneration" and "Hyperproliferative Disorders" sections below, in Example 11, 15, and 18, and elsewhere herein. Briefly, the uses include, but are not limited to the detection, treatment, and/or prevention of Alzheimer's Disease, Parkinson's Disease, Huntington's Disease, Tourette Syndrome, meningitis, encephalitis, demyelinating diseases, peripheral neuropathies, neoplasia, trauma, congenital malformations, spinal cord injuries, ischemia and infarction, aneurysms, hemorrhages, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder, depression, panic disorder, learning disabilities, ALS, psychoses, autism, and altered behaviors, including disorders in feeding, sleep patterns, balance, and perception. In addition, elevated expression of this gene product in regions of the brain indicates it plays a role in normal neural function.

Potentially, this gene product is involved in synapse formation, neurotransmission, learning, cognition, homeostasis, or neuronal differentiation or survival. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

related to SEQ ID NO:102 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or  
 5 more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1827 of SEQ ID NO:102, b is an integer of 15 to 1841, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:102, and where b is greater than or equal to a + 14.

## 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 93

The translation product of this gene shares sequence homology with a probable membrane protein YGL054c -yeast (*Saccharomyces cerevisiae*). Moreover,

The translation product of this gene also have homology to the human and mouse cornichon protein which is known to be necessary for both anterior-posterior  
 15 and dorsal-ventral pattern formation in conjunction with the EGF receptor signaling process (See Genbank Accession Nos. gblAAC98388.11 (AF104398), and spIP52159; all references and information available through these accessions are hereby incorporated herein by reference; for example, Cell 81 (6), 967-978 (1995)).

The polypeptide of this gene has been determined to have two transmembrane  
 20 domains at about amino acid position 57 - 73, and 121 - 137 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing amino acids 1 - 14 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type IIIa membrane proteins.

25 In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: YGCEKTTEGGRRRRRRMEAVVFVFSLLDCCAL  
 IFLSVYFIITLSDLECDYINARSCCSKLNKWVPELIGHTIVTVLLLMSLHWF  
 30 IFLNLPVATWNIYRYIMVPSQNMGVFDPTEIHNRGQLKSHMKEAMIKLGFH  
 LLCFFMYLYSMILALIND (SEQ ID NO:362). Polynucleotides encoding these polypeptides are also provided.

The gene encoding the disclosed cDNA is believed to reside on chromosome 1. Accordingly, polynucleotides related to this invention are useful as a marker in linkage analysis for chromosome 1.

5 This gene is expressed primarily in activated T-cells and to a lesser extent in endometrial tumor, T cell helper II cells, microvascular endothelial cells, Raji cells treated with cyclohexamide and umbilical vein endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune, hematopoietic, and vascular diseases and/or disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, vascular, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, amniotic fluid, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 216 as residues: Ser-39 to Asn-45, Asn-103 to Ser-109. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in activated T-cells indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of immune disorders involving activated T-cells. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or

other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lense tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Moreover, the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:103 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general



formula of a-b, where a is any integer between 1 to 671 of SEQ ID NO:103, b is an integer of 15 to 685, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:103, and where b is greater than or equal to a + 14.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 94

In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: ARAPAPSLPPLPSPAPALAPAHSLGLLLGRMS  
10 GSSLPSALALSLLLVSGLLPGPAAQNVRVQSGQDQ (SEQ ID NO: 363).

Polynucleotides encoding these polypeptides are also provided.

This gene is expressed primarily in dendritic cells and to a lesser extent in healing abdomen wound, and pancreas islet cell tumor cells.

Therefore, polynucleotides and polypeptides of the invention are useful as  
15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune and hematopoietic diseases and/or disorders, particularly wound healing disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential

20 identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or  
25 cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 217 as residues: Gln-34 to Lys-40. Polynucleotides  
30 encoding said polypeptides are also provided.

The tissue distribution in dendritic cells and early healing wound indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating

wounds to enhance the healing process. Representative uses are described in the "Immune Activity" and "infectious disease" sections below, in Example 11, 13, 14, 16, 18, 19, 20, and 27, and elsewhere herein. Briefly, the expression of this gene product indicates a role in regulating the proliferation; survival; differentiation; and/or

5. activation of hematopoietic cell lineages, including blood stem cells. This gene product is involved in the regulation of cytokine production, antigen presentation, or other processes suggesting a usefulness in the treatment of cancer (e.g. by boosting immune responses).

Since the gene is expressed in cells of lymphoid origin, the natural gene

10 product is involved in immune functions. Therefore it is also useful as an agent for immunological disorders including arthritis, asthma, immunodeficiency diseases such as AIDS, leukemia, rheumatoid arthritis, granulomatous Disease, inflammatory bowel disease, sepsis, acne, neutropenia, neutrophilia, psoriasis, hypersensitivities, such as T-cell mediated cytotoxicity; immune reactions to transplanted organs and

15 tissues, such as host-versus-graft and graft-versus-host diseases, or autoimmunity disorders, such as autoimmune infertility, lens tissue injury, demyelination, systemic lupus erythematosus, drug induced hemolytic anemia, rheumatoid arthritis, Sjogren's Disease, and scleroderma. Moreover, the protein may represent a secreted factor that influences the differentiation or behavior of other blood cells, or that recruits

20 hematopoietic cells to sites of injury. Thus, this gene product is thought to be useful in the expansion of stem cells and committed progenitors of various blood lineages, and in the differentiation and/or proliferation of various cell types. Furthermore, the protein may also be used to determine biological activity, raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their

25 interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are

30 related to SEQ ID NO:104 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is

cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1154 of SEQ ID NO:104, b is an integer of 15 to 1168, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:104, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 95

Contact of cells with supernatant expressing the product of this gene has been shown to increase the permeability of the plasma membrane of aortic smooth muscle cells to calcium. Thus it is likely that the product of this gene is involved in a signal transduction pathway that is initiated when the product binds a receptor on the surface of the plasma membrane of both smooth muscle cells, and in other cell-lines or tissue cell types. Thus, polynucleotides and polypeptides have uses which include, but are not limited to, activating smooth muscle cells. Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium and sodium, as well as alter pH and membrane potential. Alterations in small molecule concentration can be measured to identify supernatants which bind to receptors of a particular cell.

This gene is expressed primarily in pancreatic carcinoma, gall bladder and primary dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, metabolic and immune diseases and/or disorders, particularly cancers, such as pancreatic carcinoma and gall bladder tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., metabolic, immune, hematopoietic, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such

a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 218 as residues: Lys-34 to Ile-41. Polynucleotides  
5 encoding said polypeptides are also provided.

The tissue distribution in pancreatic carcinoma and gall bladder indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosing and treating cancer, such as pancreatic carcinoma and gall bladder tumors. Representative uses are described here and elsewhere herein. Alternatively, the  
10 detected calcium flux biological activity indicates the protein is useful in the detection, treatment, and/or prevention of a variety of vascular disorders and conditions, which include, but are not limited to microvascular disease, vascular leak syndrome, aneurysm, stroke, embolism, thrombosis, coronary artery disease, arteriosclerosis, and/or atherosclerosis. Furthermore, the protein may also be used to  
15 determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

20 Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:105 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is  
25 cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1161 of SEQ ID NO:105, b is an integer of 15 to 1175, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:105, and where b is greater than or equal to a + 14.

30

**FEATURES OF PROTEIN ENCODED BY GENE NO: 96**

The polypeptide of this gene has been determined to have a transmembrane domain at about amino acid position 10 - 26 of the amino acid sequence referenced in Table 1 for this gene. Moreover, a cytoplasmic tail encompassing amino acids 27 to 48 of this protein has also been determined. Based upon these characteristics, it is believed that the protein product of this gene shares structural features to type Ib membrane proteins.

This gene is expressed primarily in osteosarcoma, wilm's tumor, ovarian cancer and in T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, inflammatory diseases and cancers, such as osteosarcoma, wilm's tumor and ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., skeletal, renal, reproductive, immune, and cancerous and wounded tissues) or bodily fluids (e.g., lymph, serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 219 as residues: Ser-30 to Pro-35. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of inflammatory conditions and cancer, such as osteosarcoma, wilm's tumor and ovarian cancer. Moreover, the expression within cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of developmental diseases

and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern formation.

5       Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of potential roles in proliferation and differentiation, this gene product may have  
10 applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types of degenerative conditions. Thus this protein may modulate apoptosis or tissue  
15 differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new insight into the regulation of cellular growth and proliferation. Furthermore, the  
20 protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or immunotherapy targets for the above listed tissues.

25       Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:106 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is  
30 cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1007 of SEQ ID NO:106, b is an

integer of 15 to 1021, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:106, and where b is greater than or equal to a + 14.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 97

5 In another embodiment, polypeptides comprising the amino acid sequence of the open reading frame upstream of the predicted signal peptide are contemplated by the present invention. Specifically, polypeptides of the invention comprise the following amino acid sequence: GTSKDCVLYAFLDPGMAVPLFLYIFTLLPLLPF LLSLCFSPLTVKRSSSESSEKSSL (SEQ ID NO: 364). Polynucleotides encoding  
10 these polypeptides are also provided.

This gene is expressed primarily in ovarian cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
15 not limited to, ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types  
20 (e.g., reproductive, ovarian, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial amniotic fluid, fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 220 as residues: Thr-28 to Ser-40. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution in ovarian tissues indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing cancer,  
30 e.g., ovarian cancer. Moreover, the expression within cellular sources marked by proliferating cells indicates this protein may play a role in the regulation of cellular division, and may show utility in the diagnosis, treatment, and/or prevention of

developmental diseases and disorders, including cancer, and other proliferative conditions. Representative uses are described in the "Hyperproliferative Disorders" and "Regeneration" sections below and elsewhere herein. Briefly, developmental tissues rely on decisions involving cell differentiation and/or apoptosis in pattern  
5 formation.

Dysregulation of apoptosis can result in inappropriate suppression of cell death, as occurs in the development of some cancers, or in failure to control the extent of cell death, as is believed to occur in acquired immunodeficiency and certain neurodegenerative disorders, such as spinal muscular atrophy (SMA). Because of  
10 potential roles in proliferation and differentiation, this gene product may have applications in the adult for tissue regeneration and the treatment of cancers. It may also act as a morphogen to control cell and tissue type specification. Therefore, the polynucleotides and polypeptides of the present invention are useful in treating, detecting, and/or preventing said disorders and conditions, in addition to other types  
15 of degenerative conditions. Thus this protein may modulate apoptosis or tissue differentiation and is useful in the detection, treatment, and/or prevention of degenerative or proliferative conditions and diseases. The protein is useful in modulating the immune response to aberrant polypeptides, as may exist in proliferating and cancerous cells and tissues. The protein can also be used to gain new  
20 insight into the regulation of cellular growth and proliferation. Furthermore, the protein may also be used to determine biological activity, to raise antibodies, as tissue markers, to isolate cognate ligands or receptors, to identify agents that modulate their interactions, in addition to its use as a nutritional supplement. Protein, as well as, antibodies directed against the protein may show utility as a tumor marker and/or  
25 immunotherapy targets for the above listed tissues.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:107 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically  
30 excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general



formula of a-b, where a is any integer between 1 to 816 of SEQ ID NO:107, b is an integer of 15 to 830, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:107, and where b is greater than or equal to a + 14.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 98

This gene is expressed primarily in macrophages and breast cancer tissue and to a lesser extent in osteoblasts and smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune system dysfunction; inflammation; breast cancer; cancer; osteoporosis; osteopetrosis; peristaltic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and skeletal systems, expression of this gene at significantly higher or lower levels is routinely detected in certain tissues or cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid and spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Preferred polypeptides of the present invention comprise immunogenic epitopes shown in SEQ ID NO: 221 as residues: Glu-16 to Ala-40. Polynucleotides encoding said polypeptides are also provided.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and/or treatment of a variety of disorders. Expression in macrophages and other hematopoietic cell types indicates that this gene product is involved in the regulation of hematopoietic cell survival, proliferation, differentiation, or activation. It is involved in the control of such processes as immune surveillance, antigen presentation, T cell activation, cytokine release, and inflammation. Expression in breast cancer tissue may possibly correlate with the diagnosis and differentiation of cancerous tissue from normal breast tissue.

Expression in osteoblasts and osteoclasts may implicate this gene product in the process of bone turnover, and target it as a likely candidate for the treatment of osteoporosis and/or osteopetrosis. Finally, expression in smooth muscle may indicate an involvement in the normal function of numerous internal organs and in the function of the digestive system.

Many polynucleotide sequences, such as EST sequences, are publicly available and accessible through sequence databases. Some of these sequences are related to SEQ ID NO:108 and may have been publicly available prior to conception of the present invention. Preferably, such related polynucleotides are specifically excluded from the scope of the present invention. To list every related sequence is cumbersome. Accordingly, preferably excluded from the present invention are one or more polynucleotides comprising a nucleotide sequence described by the general formula of a-b, where a is any integer between 1 to 1287 of SEQ ID NO:108, b is an integer of 15 to 1301, where both a and b correspond to the positions of nucleotide residues shown in SEQ ID NO:108, and where b is greater than or equal to a + 14.

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT 3' NT of Clone Seq.	5' NT of Clone Seq.	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
1	HDPTK41	209965 06/11/98	pCMV/Sport 3.0	11	1564	1	1564	39	124	1	26	27	369
2	HFXGT26	209965 06/11/98	Lambda ZAP II	12	1757	1	1757	13	125	1	22	23	85
3	HLTGX30	209965 06/11/98	Uni-ZAP XR	13	1373	1	1373	13	126	1	41	42	43
4	HLTHG37	209965 06/11/98	Uni-ZAP XR	14	3740	1908	3740	50	127	1	1	2	319
4	HLTHG37	209965 06/11/98	Uni-ZAP XR	109	1932	98	1932	313	222	1	35	36	42
5	HNTMZ90	209965 06/11/98	pSport1	15	1196	1	1196	282	128	1	21	22	45

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
6	HPIBX03	209965 06/11/98	Uni-ZAP XR	16	2209	1	2178	81	129	1	29	30	709
7	H6EDY30	209965 06/11/98	Uni-ZAP XR	17	1774	1	1774	321	130	1	29	30	414
8	HAMGR28	209965 06/11/98	pCMV Sport 3.0	18	1674	47	1674	98	131	1	18	19	242
8	HAMGR28	209965 06/11/98	pCMV Sport 3.0	110	1534	1	1534	40	223	1	18	19	203
9	HAPNZ94	209965 06/11/98	Uni-ZAP XR	19	2018	255	2018	287	132	1	36	37	312
10	HATCP77	209965 06/11/98	Uni-ZAP XR	20	2098	1	2098	37	133	1	21	22	182

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
11	HDABR72	209965 06/11/98	pSport1	21	1746	1	1746	28	28	134	1	29	30	146
12	HDPKB18	209965 06/11/98	pCMVSPORT 3.0	22	2876	1	2876	98	98	135	1	21	22	122
12	HDPKB18	209965 06/11/98	pCMVSPORT 3.0	111	2871	1	2871	87	87	224	1	21	22	42
13	HEQCC55	209965 06/11/98	pCMVSPORT 3.0	23	1052	30	1052	62	62	136	1	27	28	112
13	HEQCC55	209965 06/11/98	pCMVSPORT 3.0	112	1037	1	1037	57	57	225	1	27	28	155
14	HETDE26	209965 06/11/98	Uni-ZAP XR	24	1541	1	1541	205	205	137	1	29	30	139

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted	Last AA of ORF
15	HOEDH84	209965 06/11/98	Uni-ZAP XR	25	2079	1	2079	256	256	138	1	20	21	404
16	HPIBT55	209965 06/11/98	Uni-ZAP XR	26	1947	129	1947	253	253	139	1	30	31	95
17	HSLCS05	209965 06/11/98	Uni-ZAP XR	27	3379	1	3354	168	168	140	1	23	24	239
18	HDPDD03	209965 06/11/98	pCMVSPORT 3.0	28	2006	1	2006	233	233	141	1	21	22	53
19	HDPDI66	209965 06/11/98	pCMVSPORT 3.0	29	3070	1	3070	93	93	142	1	45	46	66
20	HDTDQ23	209965 06/11/98	pCMVSPORT 2.0	30	2227	1	2206	148	148	143	1	20	21	108

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of 5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
20	HDTDQ23	209965 06/11/98	pCMVSPORT 2.0	113	2214	1	2206	148	148	226	1	20	21	73
21	HE2PY40	209965 06/11/98	Uni-ZAP XR	31	1288	1	1288	147	147	144	1	22	23	83
22	HEONM66	209965 06/11/98	pSPORT1	32	3280	1	3280	89	89	145	1	24	25	166
22	HEONM66	209965 06/11/98	pSPORT1	114	3300	1	3300	98	98	227	1	20	21	166
23	HKAEG43	209965 06/11/98	pCMVSPORT 2.0	33	1297	1	1297	32	32	146	1	29	30	70
23	HKAEG43	209965 06/11/98	pCMVSPORT 2.0	115	1286	1	1286	21	21	228	1	29	30	70

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted	Last AA of ORF
24	HLHDP65	209965 06/11/98	Uni-ZAP XR	34	2184	1	2184	19	19	147	1	19	20	412
24	HLHDP65	209965 06/11/98	Uni-ZAP XR	116	2189	1	2189	26	26	229	1	21	22	272
25	HLMDO03	209965 06/11/98	Uni-ZAP XR	35	949	1	949	72	72	148	1	45	46	84
26	HMAGK93	209965 06/11/98	Uni-ZAP XR	36	3338	162	1884	164	164	149	1	30	31	153
27	HMEAL02	209965 06/11/98	Lambda ZAP II	37	1563	1	1563	237	237	150	1	33	34	129
28	HMKCH52	209965 06/11/98	pSport1	38	1048	1	1048	53	53	151	1	17	18	61



Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
29	HCEFB69	209965 06/11/98	Uni-ZAP XR	39	1430	1	1430	188	188	152	1	24	25	224
30	HNFFC43	203027 06/26/98	Uni-ZAP XR	40	2103	209	2058	488	488	153	1	15	16	68
31	HSPMG77	203027 06/26/98	pSport1	41	2349	1	2349	130	130	154	1	46	47	83
32	HSQAC69	203027 06/26/98	Uni-ZAP XR	42	1559	1	1559	146	146	155	1	21	22	60
33	HSTBJ86	203027 06/26/98	Uni-ZAP XR	43	1766	1	1766	120	120	156	1	24	25	83
34	HLDQR62	203027 06/26/98	pCMVSPORT 3.0	44	2572	427	2572	520	520	157	1	18	19	161

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of 5' NT Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
35	HUVDJ43	203027 06/26/98	Uni-ZAP XR	45	526	69	526	89	89	158	1	31	32	146
36	HADCP14	203027 06/26/98	pSport1	46	1032	1	1032	35	35	159	1	20	21	142
37	HBXCF95	203027 06/26/98	ZAP Express	47	2680	1	2680	118	118	160	1	22	23	50
38	HEQBU15	203027 06/26/98	pCMVSPORT 3.0	48	1730	1	1730	56	56	161	1	26	27	64
39	HL1BD22	203027 06/26/98	Uni-ZAP XR	49	1275	1	1275	53	53	162	1	39	40	58
40	HOEEU24	203027 06/26/98	Uni-ZAP XR	50	1762	1	1762	113	113	163	1	21	22	374

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of 5' NT Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
40	HOEEU24	203027 06/26/98	Uni-ZAP XR	117	1763	1	1763	113	113	230	1	21	22	81
41	HTTBR96	203027 06/26/98	Uni-ZAP XR	51	2059	1	2059	96	96	164	1	26	27	63
42	HWHQS55	203027 06/26/98	pCMV Sport 3.0	52	3282	1	3282	169	169	165	1	26	27	742
43	HCEEK50	203027 06/26/98	Uni-ZAP XR	53	1860	1	1860	233	233	166	1	17	18	213
44	HCWBU94	203027 06/26/98	ZAP Express	54	770	1	770	109	109	167	1	26	27	212
45	HE2NR62	203027 06/26/98	Uni-ZAP XR	55	1093	1	1093	145	145	168	1	38	39	74

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of 5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
46	HHS GH19	203027 06/26/98	Uni-ZAP XR	56	632	1	632	291	291	169	1	15	16	47
47	HDP GT01	203027 06/26/98	pCMV Sport 3.0	57	2687	138	2687	8	8	170	1	28	29	87
48	HOB AF11	203027 06/26/98	pBluescript	58	619	153	579	166	166	171	1	30	31	41
49	HOH CA35	203027 06/26/98	pCMV Sport 2.0	59	1378	1	1378	153	153	172	1	15	16	47
50	HPM GP24	203027 06/26/98	Uni-ZAP XR	60	1126	1	1126	215	215	173	1	33	34	232
51	HSD IE16	203027 06/26/98	Uni-ZAP XR	61	2078	1	2078	182	182	174	1	29	30	44

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of 5' NT Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Secreted Portion	Last AA of ORF
52	HSOBK48	203027 06/26/98	Uni-ZAP XR	62	762	1	762	433	433	175	1	16	84
53	HTADH39	203027 06/26/98	Uni-ZAP XR	63	1094	1	1094	173	173	176	1	24	65
54	HUSGT36	203027 06/26/98	pSport1	64	1361	1	1361	112	112	177	1	16	54
55	HVAAE95	203027 06/26/98	pSport1	65	947	1	947	325	325	178	1	14	82
56	HHEAH25	203071 07/27/98	pCMVSPORT 3.0	66	1376	1	1376	43	43	179	1	31	330
56	HHEAH25	203071 07/27/98	pCMVSPORT 3.0	118	1375	1	1375	43	43	231	1	31	71

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
57	HBJY92	203071 07/27/98	Uni-ZAP XR	67	2434	487	2366	548	180	1	29	30	40
58	HCLCW50	203071 07/27/98	Lambda ZAP II	68	1086	1	1086	255	181	1	17	18	51
59	HDRMF68	203071 07/27/98	pSport1	69	1262	1	1262	309	182	1	22	23	54
60	HOUGG12	203071 07/27/98	Uni-ZAP XR	70	1642	35	1642	116	183	1	22	23	61
61	HEEAQ11	203071 07/27/98	Uni-ZAP XR	71	921	1	921	213	184	1	28	29	147
62	HEEAZ65	203071 07/27/98	Uni-ZAP XR	72	906	1	906	182	185	1	19	20	160

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of 5' NT Start Codon	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
63	HEGAN94	203071 07/27/98	Uni-ZAP XR	73	680	1	680	133	133	186	1	23	24	121
64	HFXBL33	203071 07/27/98	Lambda ZAP II	74	1633	1	1633	152	152	187	1	24	25	162
65	HLIBD68	203071 07/27/98	pCMVSPORT 1	75	1022	1	1022	186	186	188	1	35	36	50
66	HLTCO33	203071 07/27/98	Uni-ZAP XR	76	1184	1	1184	80	80	189	1	18	19	64
67	HLAYAC95	203071 07/27/98	pSport1	77	312	1	312	92	92	190	1	16	17	46
68	HNFGF20	203071 07/27/98	Uni-ZAP XR	78	1370	38	1370	206	206	191	1	45	46	143

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
69	HNHKS18	203071 07/27/98	Uni-ZAP XR	79	368	1	368	125	192	1	36	37	81
70	HSLJW78	203071 07/27/98	Uni-ZAP XR	80	1088	1	1088	159	193	1	20	21	44
71	HHFHD01	203071 07/27/98	Uni-ZAP XR	81	1862	1	1862	177	194	1	16	17	41
72	HLWAE11	203071 07/27/98	pCMVSPORT 3.0	82	1618	1	1618	85	195	1	27	28	259
73	HCYBN55	203071 07/27/98	pBluescript SK-	83	2034	1	1984	341	196	1	19	20	117
73	HCYBN55	203071 07/27/98	pBluescript SK-	119	1022	78	1022	3	232	1	1	2	225



Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT 3' NT of Clone Seq.	5' NT of Clone Seq.	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
74	HEONX38	203071 07/27/98	pSport1	84	2240	5	2240	23	23	1	23	24	698
74	HEONX38	203071 07/27/98	pSport1	120	2311	1	2311	24	24	1	23	24	314
75	HLDQU79	203071 07/27/98	pCMVSPORT 3.0	85	1488	1	1488	99	99	1	23	24	348
76	HSYBK21	203071 07/27/98	pCMVSPORT 3.0	86	3174	1	1466	119	119	1	29	30	401
77	HELBC12	203071 07/27/98	Uni-ZAP XR	87	2780	2110	2738	120	120	1	30	31	324
78	HTHDS25	203071 07/27/98	Uni-ZAP XR	88	1061	1	1061	70	70	1	15	16	90

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
79	HFIHO70	203071 07/27/98	pSport1	89	1342	1	1271	141	202	1	30	31	243
79	HFIHO70	203071 07/27/98	pSport1	121	1286	1	1279	131	234	1	30	31	93
80	HPMEI86	203071 07/27/98	Uni-ZAP XR	90	770	40	770	50	203	1	30	31	75
81	HSOBV29	203071 07/27/98	Uni-ZAP XR	91	1570	207	1570	244	204	1	24	25	248
82	HWABY10	203071 07/27/98	pCMVSPORT 3.0	92	2950	78	2914	263	205	1	22	23	168
83	HACCI17	203071 07/27/98	Uni-ZAP XR	93	1722	336	1714	461	206	1	24	25	218

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Start Codon	Signal Pep	AA SEQ NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
83	HACCI17	203071 07/27/98	Uni-ZAP XR	122	1380	12	1380	135	135	235	1	24	25	72
84	HAPQT22	203070 07/27/98	Uni-ZAP XR	94	635	1	635	132	132	207	1	17	18	72
85	HDPBO81	203070 07/27/98	pCMVSPORT 3.0	95	3798	1	3798	265	265	208	1	26	27	348
85	HDPBO81	203070 07/27/98	pCMVSPORT 3.0	123	3793	1	3793	255	255	236	1	26	27	348
86	HDPGI49	203070 07/27/98	pCMVSPORT 3.0	96	2683	1	2640	266	266	209	1	29	30	72
87	HDTBV77	203070 07/27/98	pCMVSPORT 2.0	97	2181	1	2181	326	326	210	1	22	23	608

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
88	HFIUE82	203070 07/27/98	pSport1	98	1957	1	1957	24	211	1	23	24	251
89	HHEND31	203070 07/27/98	pCMVSPORT 3.0	99	1112	1	1112	109	212	1	25	26	225
90	HKMND01	203069 07/27/98	pBluescript	100	887	1	887	23	213	1	26	27	50
91	HLDDB184	203069 07/27/98	pCMVSPORT 3.0	101	1248	1	1248	50	214	1	35	36	171
92	HLTEK17	203069 07/27/98	Uni-ZAP XR	102	1841	1	1841	112	215	1	13	14	47
93	HEBEJ18	203069 07/27/98	Uni-ZAP XR	103	685	7	649	51	216	1	15	16	139

Gene No.	cDNA Clone ID	ATCC Deposit Nr and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT 3' NT of Clone Seq.	5' NT of 5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	First AA of Secreted Portion	Last AA of ORF
94	HMEAI48	203069 07/27/98	Lambda ZAP II	104	1168	1 1168	95	95	217	1	29	30	40
95	HNHGN91	203069 07/27/98	Uni-ZAP XR	105	1175	161 1175	184	184	218	1	24	25	51
96	HODAE92	203069 07/27/98	Uni-ZAP XR	106	1021	1 1021	123	123	219	1	29	30	48
97	HODDF13	203069 07/27/98	Uni-ZAP XR	107	830	1 830	46	46	220	1	27	28	41
98	HCDCF30	203027 06/26/98	Uni-ZAP XR	108	1301	102 1301	151	151	221	1	14	15	40

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization

probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID  
5 NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or  
10 deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

15 Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC,  
20 as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a  
25 suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed  
30 herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

5       The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

10       The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below). It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during  
15 recombinant production.

      The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, Gene 67:31-40  
20 (1988). Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

### Signal Sequences

25       Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, Virus Res. 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, Nucleic Acids Res. 14:4683-4690 (1986) uses the  
30 information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of



these methods is in the range of 75-80%. (von Heinje, supra.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., Protein Engineering 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

### **Polynucleotide and Polypeptide Variants**

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

By a polynucleotide having a nucleotide sequence at least, for example, 95%

"identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown in Table 1, the ORF (open reading frame), or any fragment specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence

that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a query amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid.

These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

5 As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present  
10 invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity.

15 Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.

20 If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the  
25 query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from  
30 the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and

C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence.

5 For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C-termini not matched/total number of residues in the query sequence) so 10% is  
10 subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no  
15 residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for. No other manual  
20 corrections are to be made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced  
25 by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as E.  
30 coli).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an

organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level.

Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

5        Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., J. Biol. Chem. 268: 2984-2988  
10        (1993), reported variant KGF proteins having heparin binding activity even after deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

15        Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over  
20        the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in  
25        activity from wild-type.

         Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted  
30        form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic

activities can readily be determined by routine methods described herein and otherwise known in the art.

Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions,  
5 inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

10 The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these  
15 positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function.

20 For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are  
25 surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions  
30 involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of

the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention  
5 include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a  
10 substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino  
acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within  
the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of  
15 charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the  
aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev.  
20 Therapeutic Drug Carrier Systems 10:307-377 (1993).)

A further embodiment of the invention relates to a polypeptide which comprises the amino acid sequence of the present invention having an amino acid sequence which contains at least one amino acid substitution, but not more than 50  
25 amino acid substitutions, even more preferably, not more than 40 amino acid substitutions, still more preferably, not more than 30 amino acid substitutions, and still even more preferably, not more than 20 amino acid substitutions. Of course, in order of ever-increasing preference, it is highly preferable for a polypeptide to have an amino acid sequence which comprises the amino acid sequence of the present  
30 invention, which contains at least one, but not more than 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 amino acid substitutions. In specific embodiments, the number of additions, substitutions, and/or deletions in the amino acid sequence of the present invention or fragments thereof (e.g., the mature form and/or other fragments described herein), is



1-5, 5-10, 5-25, 5-50, 10-50 or 50-150, conservative amino acid substitutions are preferable.

### **Polynucleotide and Polypeptide Fragments**

5 In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt  
10 in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

15 Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-  
20 1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, or 2001 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini.  
25 Preferably, these fragments encode a polypeptide which has biological activity. More preferably, these polynucleotides can be used as probes or primers as discussed herein.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the  
30 deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the

invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

### **Epitopes & Antibodies**

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')<sub>2</sub> fragments) which are capable of specifically binding to protein. Fab and F(ab')<sub>2</sub> fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these

fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

## 5 Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second  
10 protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous  
15 functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the  
20 polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and  
25 specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-  
30 polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., Nature 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the

IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., J. Biochem. 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion  
5 proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified,  
10 would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., J. Molecular Recognition 8:52-58 (1995); K. Johanson et al., J. Biol.  
15 Chem. 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue,  
20 Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767  
25 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

### **Vectors, Host Cells, and Protein Production**

30 The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral

vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

5 The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

10 The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding  
15 portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin  
20 resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in E. coli and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as  
25 CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-  
9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and  
30 ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1

and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods  
5 are described in many standard laboratory manuals, such as Davis et al., *Basic Methods In Molecular Biology* (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

10 A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most  
15 preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical  
20 synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also  
25 include an initial modified methionine residue, in some cases as a result of host-mediated processes. Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some  
30 proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

In addition to encompassing host cells containing the vector constructs discussed herein, the invention also encompasses primary, secondary, and immortalized host cells of vertebrate origin, particularly mammalian origin, that have been engineered to delete or replace endogenous genetic material (e.g., coding sequence), and/or to include genetic material (e.g., heterologous polynucleotide sequences) that is operably associated with the polynucleotides of the invention, and which activates, alters, and/or amplifies endogenous polynucleotides. For example, techniques known in the art may be used to operably associate heterologous control regions (e.g., promoter and/or enhancer) and endogenous polynucleotide sequences via homologous recombination (see, e.g., U.S. Patent No. 5,641,670, issued June 24, 1997; International Publication No. WO 96/29411, published September 26, 1996; International Publication No. WO 94/12650, published August 4, 1994; Koller et al., Proc. Natl. Acad. Sci. USA 86:8932-8935 (1989); and Zijlstra et al., Nature 342:435-438 (1989), the disclosures of each of which are incorporated by reference in their entirety).

### **Uses of the Polynucleotides**

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids



containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library) .) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. First, visible structural alterations in the chromosomes, such as deletions or

translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991) ) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more  
5 restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

10 The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a  
15 unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological  
20 samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erlich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with  
25 one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a  
30 particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the

present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as  
5 molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an  
10 immune response.

### Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

15 A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell. Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene  
20 expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine ( $^{125}\text{I}$ ,  $^{121}\text{I}$ ), carbon ( $^{14}\text{C}$ ), sulfur ( $^{35}\text{S}$ ), tritium ( $^3\text{H}$ ), indium ( $^{112}\text{In}$ ), and technetium ( $^{99\text{m}}\text{Tc}$ ), and fluorescent labels, such as fluorescein and rhodamine, and  
25 biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which  
30 emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as

deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example,  $^{131}\text{I}$ ,  $^{112}\text{In}$ ,  $^{99\text{m}}\text{Tc}$ ), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of  $^{99\text{m}}\text{Tc}$ . The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide,  
5 such as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from  
10 a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

#### Biological Activities

15 The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.  
20

#### Immune Activity

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune  
25 cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide  
30 or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic

anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, 5 Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitis, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide 10 or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ 15 rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of 20 T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic 25 and acute conditions, including inflammation associated with infection (e.g., septic shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or 30 IL-1.)

### **Hyperproliferative Disorders**



A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenström's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

### **Infectious Disease**

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response.

Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiolitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Nocardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Meningococcal), Pasteurellaceae Infections (e.g., Actinobacillus, Haemophilus, Pasteurella), Pseudomonas, Rickettsiaceae,

Chlamydiaceae, Syphilis, and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria, Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas. These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g.,

dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

### **Regeneration**

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac), vasculature (including vascular and lymphatics), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

### **Chemotaxis**

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

### **Binding Activity**

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit (antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable

of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell  
5 membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

10 The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations,  
15 polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

20 Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers.  
25 The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

30 Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a

candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity, and (b)  
5 determining if a biological activity of the polypeptide has been altered.

### **Other Activities**

A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as  
10 discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be  
15 used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, cardiac rhythms, depression (including depressive disorders), tendency for violence,  
20 tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat  
25 content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

### **Other Preferred Embodiments**

Other preferred embodiments of the claimed invention include an isolated  
30 nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.



Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only  
5 A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for said  
10 cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the  
15 ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide  
20 sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human  
25 cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological  
30 sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X

wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing a nucleotide sequence of at least one  
5 nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid  
10 molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA  
15 molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence  
20 selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

25 The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

30 Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a

biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid  
5 sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

10 Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

15 Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

20 Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

25 Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

30 Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1

and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group

and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence  
5 selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

10 Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino  
15 acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising  
20 inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising  
25 culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of  
30 SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is

defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

### Examples

#### Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

25	<u>Vector Used to Construct Library</u>	<u>Corresponding Deposited</u>
	<u>Plasmid</u>	
	Lambda Zap	pBluescript (pBS)
	Uni-Zap XR	pBluescript (pBS)
	Zap Express	pBK
30	lafmid BA	plafmid BA
	pSport1	pSport1
	pCMVSport 2.0	pCMVSport 2.0



pCMVSPORT 3.0

pCMVSPORT 3.0

pCR<sup>®</sup>2.1pCR<sup>®</sup>2.1

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Alting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.

Vectors pSPORT1, pCMVSPORT 2.0 and pCMVSPORT 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lacmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into E. coli strain XL-1 Blue. Vector pCR<sup>®</sup>2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into E. coli strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional

plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with  $^{32}\text{P}$ - $\gamma$ -ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25  $\mu\text{l}$  of reaction mixture with 0.5  $\mu\text{g}$  of the above cDNA template. A convenient reaction mixture is 1.5-5 mM  $\text{MgCl}_2$ , 0.01% (w/v) gelatin, 20  $\mu\text{M}$  each of dATP, dCTP, dGTP, dTTP, 25

pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res. 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of

the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

**Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide**

5 A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

**Example 3: Tissue Distribution of Polypeptide**

10 Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with P<sup>32</sup> using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After  
15 labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H)  
20 or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

25

**Example 4: Chromosomal Mapping of the Polynucleotides**

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of  
30 conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing

individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

5

#### **Example 5: Bacterial Expression of a Polypeptide**

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers  
10 used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance ( $Amp^r$ ), a  
15 bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at  
20 the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance ( $Kan^r$ ). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and  
25 confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 ( $O.D.^{600}$ ) of between 0.4 and 0.6. IPTG  
30 (Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1

mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM imidazole. Imidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The

origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

#### **Example 6: Purification of a Polypeptide from an Inclusion Body**

The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without  
5 mixing for 12 hours prior to further purification steps.

To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive  
10 Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4  
15 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using  
20 a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant  $A_{280}$  monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the  
25 above refolding and purification steps. No major contaminant bands should be observed from Commassie blue stained 16% SDS-PAGE gel when 5 µg of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.



**Example 7: Cloning and Expression of a Polypeptide in a Baculovirus Expression System**

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak *Drosophila* promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.).

5       The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by  
10       gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five µg of a plasmid containing the polynucleotide is co-transfected with 1.0 µg of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method  
15       described by Felgner et al., Proc. Natl. Acad. Sci. USA 84:7413-7417 (1987). One µg of BaculoGold™ virus DNA and 5 µg of the plasmid are mixed in a sterile well of a microtiter plate containing 50 µl of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 µl Lipofectin plus 90 µl Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the  
20       transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

25       After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell  
30       culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.) After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then

resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

5 To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and  
10 cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of <sup>35</sup>S-methionine and 5 µCi <sup>35</sup>S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

15 Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

#### **Example 8: Expression of a Polypeptide in Mammalian Cells**

The polypeptide of the present invention can be expressed in a mammalian  
20 cell. A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient  
25 transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLVI, HIVI and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include,  
30 for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden), pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109),

pCMVSPORT 2.0, and pCMVSPORT 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

5           Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

10           The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker  
15 is the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991); Bebbington et al., Bio/Technology 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the  
20 production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the  
25 CMV-enhancer (Boshart et al., Cell 41:521-530 (1985).) Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

30           Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be  
5 modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

10 The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

15 Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five  $\mu$ g of the expression plasmid pC6 is cotransfected with 0.5  $\mu$ g of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded  
20 in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM,  
25 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1  $\mu$ M, 2  $\mu$ M, 5  $\mu$ M, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 -  
30 200  $\mu$ M. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

#### **Example 9: Protein Fusions**

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the half-life time in vivo. Nuclear localization signals fused to the polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:

GGGATCCGGAGCCCAAATCTTCTGACAAAACCTCACACATGCCACCGTGC  
 CCAGCACCTGAATTTCGAGGGTGCACCGTCAGTCTTCCTCTTCCCCCAAAA  
 CCAAGGACACCCTCATGATCTCCCGGACTCCTGAGGTCACATGCGTGGT  
 GGTGGACGTAAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGG  
 5 ACGGCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTA  
 CAACAGCACGTACCGTGTGGTCAGCGTCCTCACCGTCCTGCACCAGGACT  
 GGCTGAATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCA  
 ACCCCCATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAAC  
 CACAGGTGTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAG  
 10 GTCAGCCTGACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATCGCCGT  
 GGAGTGGGAGAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCT  
 CCCGTGCTGGACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTG  
 GACAAGAGCAGGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCA  
 TGAGGCTCTGCACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGG  
 15 GTAAATGAGTGCGACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

#### **Example 10: Production of an Antibody from a Polypeptide**

The antibodies of the present invention can be prepared by a variety of  
 methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a  
 20 polypeptide of the present invention is administered to an animal to induce the  
 production of sera containing polyclonal antibodies. In a preferred method, a  
 preparation of the secreted protein is prepared and purified to render it substantially  
 free of natural contaminants. Such a preparation is then introduced into an animal in  
 order to produce polyclonal antisera of greater specific activity.

25 In the most preferred method, the antibodies of the present invention are  
 monoclonal antibodies (or protein binding fragments thereof). Such monoclonal  
 antibodies can be prepared using hybridoma technology. (Köhler et al., Nature  
 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J.  
 Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell  
 30 Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures  
 involve immunizing an animal (preferably a mouse) with polypeptide or, more  
 preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in

any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

5       The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as  
10       described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

          Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method  
15       makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody  
20       whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

          It will be appreciated that Fab and F(ab')<sub>2</sub> and other fragments of the  
25       antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')<sub>2</sub> fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic  
30       chemistry.

          For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced



using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

### **Example 11: Production Of Secreted Protein For High-Throughput Screening Assays**

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at  $2 \times 10^5$  cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45

minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off PBS rinse, and person B, using a 12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of CaCl<sub>2</sub> (anhyd); 0.00130 mg/L CuSO<sub>4</sub>·5H<sub>2</sub>O; 0.050 mg/L of Fe(NO<sub>3</sub>)<sub>3</sub>·9H<sub>2</sub>O; 0.417 mg/L of FeSO<sub>4</sub>·7H<sub>2</sub>O; 311.80 mg/L of KCl; 28.64 mg/L of MgCl<sub>2</sub>; 48.84 mg/L of MgSO<sub>4</sub>; 6995.50 mg/L of NaCl; 2400.0 mg/L of NaHCO<sub>3</sub>; 62.50 mg/L of NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O; 71.02 mg/L of Na<sub>2</sub>HPO<sub>4</sub>; .4320 mg/L of ZnSO<sub>4</sub>·7H<sub>2</sub>O; .002 mg/L of Arachidonic Acid ; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of Oleic Acid; 0.010 mg/L of Palmitric Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L- Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine-H<sub>2</sub>O; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCL-H<sub>2</sub>O; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL-H<sub>2</sub>O; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-Lysine HCL; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalanine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tyrosine-2Na-2H<sub>2</sub>O; 99.65 mg/ml of L-Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319

mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B<sub>12</sub>; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine; 0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of

5 Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mm glutamine and 1x penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for

10 endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

15 On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the

20 polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

#### 25 **Example 12: Construction of GAS Reporter Construct**

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The

30 binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six

members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at  
5 higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2,  
10 Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two  
15 groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

20 Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the  
25 proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

		<u>JAKs</u>				<u>STATS</u>	<u>GAS(elements) or ISRE</u>
	<u>Ligand</u>	<u>tyk2</u>	<u>Jak1</u>	<u>Jak2</u>	<u>Jak3</u>		
<u>IFN family</u>							
5	IFN-a/B	+	+	-	-	1,2,3	ISRE
	IFN-g		+	+	-	1	GAS (IRF1>Lys6>IFP)
	Il-10	+	?	?	-	1,3	
<u>gp130 family</u>							
10	IL-6 (Pleiotrophic)	+	+	+	?	1,3	GAS (IRF1>Lys6>IFP)
	Il-11(Pleiotrophic)	?	+	?	?	1,3	
	OnM(Pleiotrophic)	?	+	+	?	1,3	
	LIF(Pleiotrophic)	?	+	+	?	1,3	
	CNTF(Pleiotrophic)	-/+	+	+	?	1,3	
15	G-CSF(Pleiotrophic)	?	+	?	?	1,3	
	IL-12(Pleiotrophic)	+	-	+	+	1,3	
<u>g-C family</u>							
	IL-2 (lymphocytes)	-	+	-	+	1,3,5	GAS
20	IL-4 (lymph/myeloid)	-	+	-	+	6	GAS (IRF1 = IFP >>Ly6)(IgH)
	IL-7 (lymphocytes)	-	+	-	+	5	GAS
	IL-9 (lymphocytes)	-	+	-	+	5	GAS
	IL-13 (lymphocyte)	-	+	?	?	6	GAS
	IL-15	?	+	?	+	5	GAS
25							
<u>gp140 family</u>							
	IL-3 (myeloid)	-	-	+	-	5	GAS (IRF1>IFP>>Ly6)
	IL-5 (myeloid)	-	-	+	-	5	GAS
	GM-CSF (myeloid)	-	-	+	-	5	GAS
30							
<u>Growth hormone family</u>							
	GH	?	-	+	-	5	

PRL	?	+/-	+	-	1,3,5	
EPO	?	-	+	-	5	GAS(B-CAS>IRF1=IFP>>Ly6)

Receptor Tyrosine Kinases

5	EGF	?	+	+	-	1,3	GAS (IRF1)
	PDGF	?	+	+	-	1,3	
	CSF-1	?	+	+	-	1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:

5':GCGCCTCGAGATTTCCCCGAAATCTAGATTTCCCCGAAATGATTTCCCC  
GAAATGATTTCCCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

5':CTCGAGATTTCCCCGAAATCTAGATTTCCCCGAAATGATTTCCCCGAAA  
TGATTTCCCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCG  
CCCCTAACTCCGCCCATCCCGCCCCCTAACTCCGCCCAGTTCCGCCCATTCT  
CCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCC  
TCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCT  
AGGCTTTTGCAAAAAGCTT:3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not  
5 contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance  
10 gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS  
15 with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, IL-  
20 2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

**Example 13: High-Throughput Screening Assay for T-cell Activity.**

25 The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells  
30 (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.



Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1% Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies) with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells ( $10^7$  per transfection), and resuspend in OPTI-MEM to a final concentration of  $10^7$  cells/ml. Then add 1ml of  $1 \times 10^7$  cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Gentamicin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100,000 cells per well).

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the  
5 assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophene covers) and  
10 stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the  
15 positive control wells.

The above protocol may be used in the generation of both transient, as well as, stable transfected cells, which would be apparent to those of skill in the art.

#### **Example 14: High-Throughput Screening Assay Identifying Myeloid Activity**

20 The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells. Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell  
25 used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest  $2 \times 10^7$  U937 cells and  
30 wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing

10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM  
5 KCl, 375 uM  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ , 1 mM  $\text{MgCl}_2$ , and 675 uM  $\text{CaCl}_2$ . Incubate at 37°C for 45 min.

Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400  
10 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

These cells are tested by harvesting  $1 \times 10^8$  cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of  $5 \times 10^5$  cells/ml. Plate 200 ul cells per well in  
15 the 96-well plate (or  $1 \times 10^5$  cells/well).

Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant  
20 according to the protocol described in Example 17.

#### **Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.**

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes,  
25 EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat pheochromocytoma cells) are known to proliferate and/or  
30 differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor).

The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

5 The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)

5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

10 Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

15 To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

20 PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

25 Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

30 To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS

(Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count  
5 the cell number and add more low serum medium to reach final cell density as  $5 \times 10^5$  cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to  $1 \times 10^5$  cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR  
10 can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

#### **Example 16: High-Throughput Screening Assay for T-cell Activity**

15 NF- $\kappa$ B (Nuclear Factor  $\kappa$ B) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- $\kappa$ B regulates the expression of genes involved in immune cell activation, control of  
20 apoptosis (NF-  $\kappa$ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF-  $\kappa$ B is retained in the cytoplasm with I- $\kappa$ B (Inhibitor  $\kappa$ B). However, upon stimulation, I-  $\kappa$ B is phosphorylated and degraded, causing NF-  $\kappa$ B to shuttle to the nucleus, thereby activating transcription of target  
25 genes. Target genes activated by NF-  $\kappa$ B include IL-2, IL-6, GM-CSF, ICAM-1 and class I MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- $\kappa$ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- $\kappa$ B would be useful in  
30 treating diseases. For example, inhibitors of NF- $\kappa$ B could be used to treat those

diseases related to the acute or chronic activation of NF- $\kappa$ B, such as rheumatoid arthritis.

To construct a vector containing the NF- $\kappa$ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- $\kappa$ B binding site (GGGGACTTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:  
5':GCGGCCTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGAC  
TTTCCATCCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:

5':GCGGCAAGCTTTTTGCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene)

Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5':CTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGGACTTTCC  
ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCC  
ATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGA  
CTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTA  
TTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAA  
GCTT:3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2-promoter plasmid (Clontech) with this NF- $\kappa$ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- $\kappa$ B/SV40/SEAP cassette is removed from the above NF- $\kappa$ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly,

the NF- $\kappa$ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF- $\kappa$ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

#### **Example 17: Assay for SEAP Activity**

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15  $\mu$ l of 2.5x dilution buffer into Optiplates containing 35  $\mu$ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50  $\mu$ l Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50  $\mu$ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

#### **Reaction Buffer Formulation:**

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5

13	75	3.75
14	80	4
15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6
23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

---



**Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability**

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-4 (Molecular Probes, Inc.; catalog no. F-14202), used here.

For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO<sub>2</sub> incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-4 is made in 10% pluronic acid DMSO. To load the cells with fluo-4, 50 ul of 12 ug/ml fluo-4 is added to each well. The plate is incubated at 37°C in a CO<sub>2</sub> incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10<sup>6</sup> cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-4 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10<sup>6</sup> cells/ml, and dispensed into a microplate, 100 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-4. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular signaling event which has resulted in an increase in the intracellular  $\text{Ca}^{++}$  concentration.

#### Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase (RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford, MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford, MA) are used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na<sub>3</sub>VO<sub>4</sub>, 2 mM Na<sub>4</sub>P<sub>2</sub>O<sub>7</sub> and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

- Generally, the tyrosine kinase activity of a supernatant is evaluated by
- 5 determining its ability to phosphorylate a tyrosine residue on a specific substrate (a biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.
- 10 The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg<sub>2+</sub> (5mM ATP/50mM MgCl<sub>2</sub>), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl<sub>2</sub>, 5 mM MnCl<sub>2</sub>, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the
- 15 components gently and preincubate the reaction mix at 30°C for 2 min. Initiate the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mM EDTA and place the reactions on ice.

- Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction
- 20 mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as
- 25 above.

- Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of
- 30 tyrosine kinase activity.

**Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity**

As a potential alternative and/or complement to the assay of protein tyrosine kinase activity described in Example 19, an assay which detects activation  
5 (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase, Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other  
10 phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G  
15 plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1

and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at  
20 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts  
25 filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place  
of A431 extract. Plates are then treated with a commercial polyclonal (rabbit)  
30 antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive

incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

5    **Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide**

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA  
10    is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products are then sequenced using primers labeled at their 5' end with T4  
15    polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

20    PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining  
25    alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenindeoxy-uridine 5'-triphosphate (Boehringer Mannheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to  
30    the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for

precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and  
5 chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

10  
**Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample**

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide  
15 is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with  
20 specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample  
25 containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature.  
30 The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

- Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale).  
5 Interpolate the concentration of the polypeptide in the sample using the standard curve.

**Example 23: Formulating a Polypeptide**

- 10 The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective  
15 amount" for purposes herein is thus determined by such considerations.

- As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1  $\mu\text{g/kg/day}$  to 10  $\text{mg/kg/day}$  of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01  
20  $\text{mg/kg/day}$ , and most preferably for humans between about 0.01 and 1  $\text{mg/kg/day}$  for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1  $\mu\text{g/kg/hour}$  to about 50  $\mu\text{g/kg/hour}$ , either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of  
25 treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

- Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal  
30 patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to



modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules. Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D-(-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's

solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in

the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other  
5 therapeutic compounds.

**Example 24: Method of Treating Decreased Levels of the Polypeptide**

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by  
10 administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

15 For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

---

20 **Example 25: Method of Treating Increased Levels of the Polypeptide**

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

25 For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

30

**Example 26: Method of Treatment Using Gene Therapy**

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

#### **Example 27: Method of Treatment Using Gene Therapy - In Vivo**

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide. The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, WO90/11092, WO98/11779; U.S. Patent NO. 5693622, 5705151, 5580859; Tabata H. et al. (1997) Cardiovasc. Res. 35(3):470-479, Chao J et al. (1997) Pharmacol. Res. 35(6):517-522, Wolff J.A. (1997) Neuromuscul. Disord. 7(5):314-318, Schwartz B. et al. (1996) Gene Ther. 3(5):405-411, Tsurumi Y. et al. (1996) Circulation 94(12):3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The

polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P.L. et al. (1995) Ann. NY Acad. Sci. 772:126-139 and Abdallah B. et al. (1995) Biol. Cell 85(1):1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely

differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 g/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant-DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e.g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A

time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

- 5 The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

**Example 28: Transgenic Animals.**

- 10 The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, *e.g.*, baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are  
15 used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

- Any technique known in the art may be used to introduce the transgene (*i.e.*, polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear  
20 microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40:691-698 (1994); Carver et al., Biotechnology (NY) 11:1263-1270 (1993); Wright et al., Biotechnology (NY) 9:830-834 (1991); and Hoppe et al., U.S. Pat. No. 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82:6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic  
25 stem cells (Thompson et al., Cell 56:313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell. Biol. 3:1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, *e.g.*, Ulmer et al., Science 259:1745 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-  
30 mediated gene transfer (Lavitrano et al., Cell 57:717-723 (1989); etc. For a review of such techniques, see Gordon, "Transgenic Animals," Intl. Rev. Cytol. 115:171-229 (1989), which is incorporated by reference herein in its entirety.



Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., Nature 380:64-66 (1996); Wilmut et al., Nature 385:810-813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i.e.*, mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, *e.g.*, head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89:6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265:103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse

transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

#### **Example 29: Knock-Out Animals.**

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (*E.g.*, see Smithies et al., Nature 317:230-234 (1985); Thomas & Capecchi, Cell 51:503-512 (1987); Thompson et al., Cell 5:313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in

the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (*e.g.*, see Thomas & Capecchi 1987 and Thompson 1989, *supra*). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

10 In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (*i.e.*, animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, ~~*e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the~~ transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc. The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U.S. Patent No. 5,399,349; and

Mulligan & Wilson, U.S. Patent No. 5,460,959 each of which is incorporated by reference herein in its entirety).

5 When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

10 Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

15 It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

20 The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference. Further, the hard copy of the sequence listing submitted herewith and the corresponding computer readable form are both incorporated herein by reference in their entireties.

25

Applicant's or agent's file reference number	PZ031PCT	International application No.	Unassigned
---	----------	-------------------------------	------------

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

<b>A.</b> The indications made below relate to the microorganism referred to in the description on page <u>259</u> , line <u>N/A</u>	
<b>B. IDENTIFICATION OF DEPOSIT</b> Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit July 27, 1998	Accession Number 203070
<b>C. ADDITIONAL INDICATIONS</b> (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
<b>D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE</b> (if the indications are not for all designated States)	
<b>E. SEPARATE FURNISHING OF INDICATIONS</b> (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	

For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer	Yvette E. Simms PCT International Division

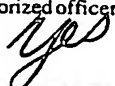
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

Applicant's or agent's file reference number	PZ031PCT	International application No.	Unassigned
---	----------	-------------------------------	------------

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>260</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT <span style="float: right;">Further deposits are identified on an additional sheet <input type="checkbox"/></span>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit July 27, 1998	Accession Number 203069
C. ADDITIONAL INDICATIONS (leave blank if not applicable) <span style="float: right;">This information is continued on an additional sheet <input type="checkbox"/></span>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g. "Accession Number of Deposit")	

<input checked="" type="checkbox"/> For receiving Office use only This sheet was received with the international application
Authorized officer  Yvonne E. Simms PCT International Division

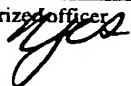
<input type="checkbox"/> For International Bureau use only This sheet was received by the International Bureau on:
Authorized officer

Applicant's or agent's file reference number	PZ031PCT	International application No.	Unassigned
---	----------	-------------------------------	------------

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>253</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 10801 University Boulevard Manassas, Virginia 20110-2209 United States of America	
Date of deposit July 27, 1998	Accession Number 203071
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	

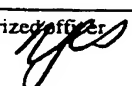
<input checked="" type="checkbox"/> For receiving Office use only This sheet was received with the international application	<input type="checkbox"/> For International Bureau use only This sheet was received by the International Bureau on:
Authorized officer  Yvette E. Simms PCT International Division	Authorized officer

Applicant's or agent's file reference number	PZ031PCT	International application No.	Unassigned
---	----------	-------------------------------	------------

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

<b>A.</b> The indications made below relate to the microorganism referred to in the description on page <u>249</u> , line <u>N/A</u>	
<b>B. IDENTIFICATION OF DEPOSIT</b> Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution <u>American Type Culture Collection</u>	
Address of depositary institution (including postal code and country) <u>10801 University Boulevard</u> <u>Manassas, Virginia 20110-2209</u> <u>United States of America</u>	
Date of deposit <u>June 26, 1998</u>	Accession Number <u>203027</u>
<b>C. ADDITIONAL INDICATIONS</b> (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
<b>D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE</b> (if the indications are not for all designated States)	
<b>E. SEPARATE FURNISHING OF INDICATIONS</b> (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")          	

<input checked="" type="checkbox"/> For receiving Office use only
<input checked="" type="checkbox"/> This sheet was received with the international application.
Authorized officer  <b>Yvette E. Simms</b> PCT International Division


<input type="checkbox"/> For International Bureau use only
<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer



Applicant's or agent's file reference number	PZ031PCT	International application No.	Unassigned
---	----------	-------------------------------	------------

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page <u>243</u> , line <u>N/A</u>	
B. IDENTIFICATION OF DEPOSIT <span style="float: right;">Further deposits are identified on an additional sheet <input type="checkbox"/></span>	
Name of depositary institution <u>American Type Culture Collection</u>	
Address of depositary institution (including postal code and country) <u>10801 University Boulevard</u> <u>Manassas, Virginia 20110-2209</u> <u>United States of America</u>	
Date of deposit <u>June 11, 1998</u>	Accession Number <u>209965</u>
C. ADDITIONAL INDICATIONS (leave blank if not applicable) <span style="float: right;">This information is continued on an additional sheet <input type="checkbox"/></span>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")	
<input checked="" type="checkbox"/> For receiving Office use only This sheet was received with the international application  Authorized officer  <b>Yvette E. Simms</b> PCT International Division	<input type="checkbox"/> For International Bureau use only This sheet was received by the International Bureau on:  Authorized officer

Form PCT/RO/134 (July 1992)

*What Is Claimed Is:*

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(c) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(d) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;

(e) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X, having biological activity;

(f) a polynucleotide which is a variant of SEQ ID NO:X;

(g) a polynucleotide which is an allelic variant of SEQ ID NO:X;

(h) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y;

(i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.

2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein.

3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.
4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.
5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.
7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.
8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.
9. A recombinant host cell produced by the method of claim 8.
10. The recombinant host cell of claim 9 comprising vector sequences:
11. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:

- (a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z, having biological activity;
  - (c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (e) a secreted form of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (f) a full length protein of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;
  - (g) a variant of SEQ ID NO:Y;
  - (h) an allelic variant of SEQ ID NO:Y; or
  - (i) a species homologue of the SEQ ID NO:Y.
12. The isolated polypeptide of claim 11, wherein the secreted form or the full length protein comprises sequential amino acid deletions from either the C-terminus or the N-terminus.
13. An isolated antibody that binds specifically to the isolated polypeptide of claim 11.
14. A recombinant host cell that expresses the isolated polypeptide of claim 11.
15. A method of making an isolated polypeptide comprising:
- (a) culturing the recombinant host cell of claim 14 under conditions such that said polypeptide is expressed; and
  - (b) recovering said polypeptide.
16. The polypeptide produced by claim 15.

17. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11 or the polynucleotide of claim 1.

18. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

19. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:

(a) contacting the polypeptide of claim 11 with a binding partner; and

(b) determining whether the binding partner effects an activity of the polypeptide.

21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.

22. A method of identifying an activity in a biological assay, wherein the method comprises:

(a) expressing SEQ ID NO:X in a cell;

(b) isolating the supernatant;

(c) detecting an activity in a biological assay; and

(d) identifying the protein in the supernatant having the activity.

23. The product produced by the method of claim 20.

<110> Human Genome Sciences, Inc.

<120> 98 Human Secreted Proteins

<130> PZ031.PCT

<140> Unassigned

<141> 1999-07-28

<150> 60/094,657

<151> 1998-07-30

<150> 60/095,486

<151> 1998-08-05

<150> 60/095,455

<151> 1998-08-06

<150> 60/095,454

<151> 1998-08-06

<150> 60/096,319

<151> 1998-08-12

<160> 364

<170> PatentIn Ver. 2.0

<210> 1

<211> 733

<212> DNA

<213> Homo sapiens

<400> 1

gggatccgga	gcccaaattct	tctgacaaaa	ctcacacatg	cccaccgtgc	ccagcacctg	60
aattcgaggg	tgcaccgtca	gtcttctctt	tccccccaaa	acccaaggac	accctcatga	120
tctcccggac	tcttgaggtc	acatgcgtgg	tggtggacgt	aagccacgaa	gaccctgagg	180
tcaagttcaa	ctggtacgtg	gacggcgtgg	aggtgcataa	tgccaagaca	aagccgcggg	240
aggagcagta	caacagcacg	taccgtgtgg	tcagcgtcct	caccgtcctg	caccaggact	300
ggctgaatgg	caaggagtac	aagtgcaagg	tctccaaāaa	agccctccca	accccatcg	360
agaaaaccat	ctccaaagcc	aaagggcagc	cccgagaacc	acaggtgtac	accctgcccc	420
catcccggga	tgagctgacc	aagaaccagg	tcagcctgac	ctgcctgggc	aaaggcttct	480
atccaagcga	catcgccgtg	gagtgggaga	gcaatgggca	gccgggagac	aactacaaga	540
ccacgcctcc	cgtgctggac	tccgacggct	ccttcttctt	ctacagcaag	ctcaccgtgg	600
acaagagcag	gtggcagcag	gggaacgtct	tctcatgctc	cgtgatgcat	gaggctctgc	660
acaaccacta	cacgcagaag	agcctctccc	tgtctccggg	taaatgagtg	cgacggccgc	720
gactctagag	gat					733

<210> 2

<211> 5

<212> PRT

<213> Homo sapiens

<220>

<221> Site

<222> (3)

<223> Xaa equals any of the twenty naturally occurring L-amino acids

<400> 2

Trp Ser Xaa Trp Ser  
1 5

<210> 3

<211> 86

<212> DNA

<213> Homo sapiens

<400> 3

gcgcctcgag atttccccga aatctagatt tccccgaaat gatttccccg aaatgatttc 60  
cccgaatat ctgcatctc aattag 86

<210> 4

<211> 27

<212> DNA

<213> Homo sapiens

<400> 4

gcggcaagct ttttgcaaag cctaggc 27

<210> 5

<211> 271

<212> DNA

<213> Homo sapiens

<400> 5

ctcgagattt ccccgaaatc tagatttccc cgaaatgatt tccccgaaat gatttccccg 60  
aaatatctgc catctcaatt agtcagcaac catagtcccc cccctaactc cgcccatccc 120  
gcccctaact ccgcccagtt ccgcccattc tccgccccat ggctgactaa ttttttttat 180  
ttatgcagag gccgaggccg cctcggcctc tgagctattc cagaagtagt gaggaggctt 240  
ttttggaggc ctaggctttt gcaaaaagct t 271

<210> 6

<211> 32

<212> DNA

<213> Homo sapiens

<400> 6

gcgctcgagg gatgacagcg atagaacccc gg 32

<210> 7

<211> 31

<212> DNA

<213> Homo sapiens

<400> 7

gcgaagcttc gcgactcccc ggatccgcct c 31

<210> 8

<211> 12

<212> DNA



&lt;213&gt; Homo sapiens

&lt;400&gt; 8

ggggactttc cc

12

&lt;210&gt; 9

&lt;211&gt; 73

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 9

gcggcctcga ggggactttc ccggggactt tccggggact ttccgggact ttccatcctg  
ccatctcaat tag

60

73

&lt;210&gt; 10

&lt;211&gt; 256

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 10

ctcgagggga ctttcccgga gactttccgg ggactttccg ggactttcca tctgccatct  
caattagtcg gcaaccatag tcccgccctt aactccgccc atcccgcccc taactccgcc  
cagttccgcc cattctccgc cccatggctg actaattttt tttatttatg cagaggccga  
ggccgcctcg gcctctgagc tattccagaa gtagtgagga ggcttttttg gaggcctagg  
cttttgcaaa aagctt

60

120

180

240

256

&lt;210&gt; 11

&lt;211&gt; 1564

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 11

gcggacgggtg ggctgtgcaa ccttcctccc tttctttaa	gcttggggca tttgtctg	gc
cttccctttt actgctggct gctgcctgca tctgtctctt	aaccttcatt aactgtgcct	120
atgtcaaatg ggaaccctg gtacaagata ttttcaccta	tgctaaagta ttggcactga	180
tcgcggtcat cgttgcaggc attgttagac ttggccaggg	agcctctact cattttgaga	240
attcctttga gggttcatca tttgcagtgg gtgacattgc	cctggcactg tactcagctc	300
tggtctccta ctcaggctgg gacacctca actatgtcac	tgaagagatc aagaatcctg	360
agaggaacct gcccctctcc attggcatct ccatgccc	atctatatct	420
tgaccaatgt ggctattat actgtgctag acatgagaga	catcttggcc agtgatgctg	480
ttgctgtgac ttttgagat cagatatttg gaatatttaa	ctggataatt ccaactgtcag	540
ttgcattatc ctgttttggg ggctcaatg cctccattgt	ggctgcttct aggttttct	600
ttgtgggctc aagagaaggc catctccctg atgccatctg	catgatccat gttgagcggg	660
tcacaccagt gccttctctg ctcttcaatg gtatcatggc	attgatctac ttgtgcgtgg	720
aagacatctt ccagctcatt aactactaca gcttcagcta	ctggttcttt gtggggcttt	780
ctattgtggg tcagctttat ctgcgctgga aggagcctga	tcgacctcgt cccctcaagc	840
tcagcgtttt cttcccgatt gtcttctgcc tctgcacct	cttccctggg gctgttccac	900
tttacagtga tactatcaac tccctcatcg gcattgccat	tgccctctca ggctgcct	960
tttacttctt catcatcaga gtgccagaac ataagcgacc	gctttacctc cgaagatcgt	1020
ggggtctgcc acaaggatcc tccaggctcct gtgtatgtca	gttgctgcag aaatggattt	1080
ggaagatgga ggagagatgc ccaagcaacg ggatcccaag	tctaactaaa caccatctgg	1140
aatcctgatg tggaaagcag gggtttcttg tctactggct	agagctaagg aagttgaaaa	1200
ggaaagctca cttctttgga ggcacctgtc cagaagcctg	gcctaggcag cttcaacctt	1260
tgaacttact ttttgaaatg aaaagtaatt tatttgtttt	gctacatact gttccagact	1320
tttaaagggg acaatgaagg tgactgtggg gaggagcatg	tcagggtttg gcttggttgt	1380
tttagaagca cctgggtgtg cctacctact cctcttttct	tttaaaaggg cccacaatgc	1440

tccaattttcc	tgtctccttt	agagagacat	gaaactatca	cagggtgctgg	atgacaataa	1500
aagttttatgt	tcctaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaagggcgcg	1560
cccg						1564

<210> 12  
 <211> 1757  
 <212> DNA  
 <213> Homo sapiens

<400> 12						
ggcacgaggt	agatggggtt	ttgggtgtga	tgtcctttct	gtttgttagt	tgtccttcta	60
acagacagga	ccctcagctg	cagggtctgt	ggagtacct	gcaatgtgag	gtgtcagtgt	120
gcccctgctg	gaggggtgct	cccagttagg	ctgctcgcg	gtcaggggtc	agggaccac	180
ttgaggaggc	agtctgccc	ttctcagatc	tccagctgca	tgtctgggaga	accactgctc	240
tcttcaaagc	tgtcagacag	ggacatttaa	gtctgcagag	gttactgctg	tctttttgtt	300
tgtctgtgcc	ctgccccag	aggtggagcc	tacagaggca	ggcaggcctc	cttgagctgt	360
gggtgggctcc	acccagttcg	agcttcccgg	ctgctttgtt	tacctaaagca	agcctgggca	420
atggcgggcg	ccccctcccc	agtctcgctg	ccgccttgca	gtttgatctc	agactgctgt	480
gctagcaatc	agcgagactc	cgtggggtag	gaccctctga	gccagggtgg	ggatataatc	540
ttgtggtgcy	ccatttttta	agcccgtcgg	aaaagcgcag	tattcagggt	ggagtgaccc	600
gattttcccg	attttccagg	tgcgcgtcgt	catcccttct	tttgattagg	aaaggggaact	660
ccctgacccc	ttgcgcttcc	cgagtgaggc	aatgccttgc	gctgctttgg	ctcgtgcatg	720
gtgcgtgcac	ccactgacct	gcgcccactg	tctggcactc	cctagtgaga	tgaacccggt	780
acctcagatg	gaaatgcaga	aatcacccgt	cttctgcgtt	gctcaggctg	ggagctgtag	840
accagagctg	ttcctattcc	gccatcttgg	ctcctcccc	accccccgcc	caatcctaag	900
tttatctctt	aacaaagtaa	atactgcaca	cgtattttaca	tgttttaaaa	tattgtttat	960
ggagcactta	ctatgtgttt	tacggtccaa	gtactttata	ttattcactg	actctttata	1020
gcaactctat	aacatattaa	tagattgtct	caatttcact	gatgacaaaa	ctggcagctt	1080
agtaactaat	gagtaacaat	accagcccct	aaaacctagg	gtctttccac	tacagcacac	1140
tctgactgac	attgctcttg	ttcagacaca	gcactcttca	cctagaaaag	gatggagagc	1200
tagaaaagtg	ttttcttttt	caagttgttc	ccattcatca	atattctatc	aagatatatg	1260
gaagagttat	tttctaggtc	ttccagaagg	ggctgctgac	acttgggaat	tcaaaaataat	1320
tactccagag	ggaaatgtgg	gcatagaaac	tttttaaaaa	ctgaattttc	ccgatctaata	1380
atcggtcttc	acatacagta	tattatttct	taaaaatcag	ccagaaatgc	acctgcagtt	1440
ttatctactg	gattttctct	tcccagcccc	attcttcttt	aagtagtttt	ttttcccat	1500
tttataaaag	aaaagcatac	ctcccaccta	tagcttccct	ccttatggta	tattatctca	1560
ggggataatt	tgaaatattg	cttgagacac	aattctcatt	aagaaataaa	ctttcaggct	1620
gggtgcagtg	gctcacactt	gtaatccccg	aactttggga	ggccaaggcg	ggggaggatc	1680
cctgaaggcc	aagagttcaa	gaccagcctg	gaatatatac	caagatcccc	tctccacaaa	1740
aaaaaaaaaa	aaaaaaa					1757

<210> 13  
 <211> 1373  
 <212> DNA  
 <213> Homo sapiens

<400> 13						
gtgaaatgaa	atatgtatac	caaaacacat	aaattttaagt	tttataaatt	tctttccttg	60
tggatctgga	aaattttctt	tctgctgttt	ttcattctga	tagttgcttt	ggctttccct	120
attccctgcc	tttccatatt	ttgaaatgtt	ttgggtcaaat	tttttctgt	tgagtgtgca	180
gtatgttttg	attgttaatt	ttattgacca	tcattctcagt	ggattattta	tttgaaactt	240
ttctgtttcc	tgttctggaa	tgtttctctg	tctggaagcc	atatttttaac	tgtatactga	300
tccagcttta	tttatcttaa	gtatttttac	aattatttta	ctgataatat	ataactaggt	360
ataatcactc	tagtcagtga	aatcaaaactg	ataacacaaa	atgatgctgc	tagtaaatct	420
agcctgaaaa	ggatcagagg	aagcactttt	tagtctaagt	agtcaagatc	aataaaactta	480
gaatgagagt	tttgtattat	ttagccttgt	gatagaagga	aaaagtacag	ttattgaatc	540
tgtatattaa	tttatgcaaa	atacacacat	tatctttcca	ttaacacagt	agtatgacat	600

acagttttat	caccactgga	ttcatgtatg	cttaaaatga	aaatgaaagg	aacacacagg	660
aaagtagctt	cttatcttcc	cagatacagc	ctccaagagt	gcctctgctg	tgtactggag	720
tgagtttgaa	gaagttcagg	taatcctctc	ctaaccagca	gcttaactta	aagtgcattc	780
tcatttttgt	cctaccacta	aaagcttgct	tattgacctc	cttctttcct	gtctcctaaa	840
ttgtgactgg	ctctcagtta	gacttgcatc	tggttgctatt	tgtagcagca	attgcaggat	900
tcatttgctt	tttctcttgc	agtgtatttc	tgaatgtttt	ataactctgg	gtatgcataa	960
tgtacgtata	tgcatatcca	tatggccaag	agtttttcaga	attcttttgt	tgaataatga	1020
atgtgtgtat	tccctacccc	ttgttttgct	ctgtcttaac	atttaagctt	tgcttattaa	1080
aatgtatttg	ttttaatgct	ttctctttag	attgtaagct	cttcatgggc	agtggtcaga	1140
tattctgtat	ttgtaacttc	ccgagaatct	agtattttat	catgtacata	atgaacatag	1200
aatagtattg	attggaagga	cagtgaagta	caaaatgaga	ttctaaaatg	acattttcat	1260
tagggctcac	aagtgtataa	gaagtatccc	tttcttattt	ttttgtttaa	tactttagaa	1320
gctattacat	taaaaaagct	attctcatta	aaaaaaaaaa	aaaaaaaaaa	aaa	1373

&lt;210&gt; 14

&lt;211&gt; 3740

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (957)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 14

actctctaag	gctttttttg	actcaccaaa	caggcttctt	gcagtgagga	tgaatacaga	60
tcacttaagg	ttgactgtgc	caaattggcat	aggggccctg	aagctaaggg	aaatggaaca	120
ctactttctca	cagggcctgt	cagttcagct	gtttaatgat	gggtccaagg	gcaaactcaa	180
tcattttatgt	ggagctgact	ttgtgaaaag	tcatacagaa	cctccacagg	gaatggaaat	240
taagtccaat	gaaagatgct	gttcttttga	tgagatgca	gacagaattg	tttattacta	300
ccatgatgca	gatggccact	ttcatctcat	agatggagac	aagatagcaa	cgtaatttag	360
cagtttcctt	aaagagctcc	tggtggagat	tgagaaaagt	ttgaatattg	gtgtgtgaca	420
aactgcata	gcaaatggaa	gttcaacacg	gtatcttgaa	gaagtattga	aggtacctgt	480
ctattgcact	aagactgggtg	taaaacattt	gcaccacaag	gctcaagagt	ttgacattgg	540
agttttatttt	gaagcaaatg	ggcatggcac	tgcaactgtt	agtacagctg	ttgaaatgaa	600
gataaaacaa	tcagcagaac	aactggaaga	taagaaaaga	aaagctgcta	agatgcttga	660
aaacattatt	gacttggtta	accaggcagc	tggtgatgct	atttctgaca	tgctgggtgat	720
tgaagcaatc	ttggctctga	agggtctgac	tgtacaacag	tgggatgctc	tctatacaga	780
tcttccaaac	agacaactta	aagttcaggt	tgcagacagg	agagttatta	gcactaccra	840
tgctgaaaga	caagcagtta	cacccccagg	attacaggag	gcaatcaatg	acctggtgaa	900
gaagtacaag	ctttctcgag	cttttgtccg	gccctctggt	acagaagatg	tcgtccngag	960
tatatgcaga	agcagactca	caagaaaagt	cagatcacct	tgacatgaa	gtgagcttgg	1020
cagtattttca	gctggctgga	ggaattggag	aaaggcccca	accagggtttc	tgaagataat	1080
tttcatattc	ctgagaaact	ggacttttta	caagtcttta	caaaactgtc	aataataatg	1140
gcagtactaa	gagattttata	atcataatgt	ttacaatgca	gcctactgga	ttgtctctag	1200
atctgttttt	cttaaacact	aacagaataa	ttcttttata	ataggtaagc	cttacacttg	1260
ttaaagaaat	ttacctctaa	tttcagctc	actaatgtaa	aatactggga	cttaagtata	1320
caattcagtc	actaactgta	cagttttatg	tgggggaacaa	ttcatgcagg	ctactggaaa	1380
attaaatctt	attaccaact	ccttgtgata	tctttgccat	caccatcaca	tgagcaagat	1440
gatgttttgc	agcattcccc	attgctgata	caaattggaga	gggcagagaa	gactttatac	1500
aaccagtttt	tccattgcag	agtcttaaga	aagattatta	gatgacttac	ctatatgact	1560
aatgccatca	ggaactcaga	ggtatgaata	gggggttgct	catccctctt	ccatactgag	1620
gtggagatgc	tcatgcaata	cttttaagga	tgcatggctc	agccttcagt	tattcttcac	1680
tgctcttggt	gaaggtatgt	gggagaaaaa	ctaattataa	tacgtttccc	agcctctgat	1740
ggagaaggaa	caccattctg	ataccagaac	atggtttaata	aggaaaagag	aaaaatcccc	1800
aaccaatctt	aattgaacca	agtctgaaac	caatggaaaa	aaaaaatggg	tagtgtatat	1860
tttgagggtt	taagacaact	caggacaata	aaaacaatgg	actttacatg	tgtatatata	1920
tagctctctt	aggcaccata	atcagtatga	gccaacaata	tttaaacttg	attcaggcca	1980

cattcagaca	tttgccttta	tatacaaaata	tttaaattaa	atacaatctg	aaatgtgttc	2040
tggtacatac	aaaaaaggaa	aaactataca	acgcagagca	gtgtgtgtgt	tttaaataat	2100
tacattttaca	tgtaagctaa	atggaaccag	caatggtgct	caagttttta	tcattcccttc	2160
cagaaaaatct	ttttctacca	tctcttctat	tttttgctg	gctttgctgg	aacatggttt	2220
gtggttctcc	agtttcatgt	ccttattagg	gaaggcattt	gagtagagga	taggactccc	2280
tgagtgtcct	ccacatcggc	ttgtgacttt	gctgttgaag	acttgactga	gcacattgaa	2340
gaacggcagg	agctgctcca	tactgcgac	ggtgcagatg	gtgagcagca	agtgccttgg	2400
ctcccaaccc	aatgttctcc	ctgagttgtc	ttcctctgga	ttttctgca	gaaaacaaaa	2460
agtgaactgg	tattaataca	acagacaatg	tggtatgtta	gaaaaattaa	aaatatataa	2520
actttggcaa	ttggtcaaga	aatgaatata	aatgacatta	agtttctaac	tcctgacctg	2580
atcaaaaacc	ttggtgcttc	tgagaccttt	tactgccatt	tattagtttt	acatggagca	2640
gtctaacatt	gtagtaatag	ttcccaacta	gaatgcgag	ataagcttag	ttaacagaaa	2700
tagctttgaa	caggaataga	gtcaaacata	aaagttttat	gttgtgcttt	gtattttactc	2760
aaaaagctcc	caggtttctg	aaccctcact	actgtaacca	aggactaggt	cacaaaatta	2820
ctacagaaaa	aaggaaacaaa	gtgctttata	catttcataa	tatatccctt	tttattataa	2880
ttagttaatt	cccttttatc	taaatggcct	aaatttgcca	tgatggtagc	agtgtccaaa	2940
gtgaataatt	actgtcagta	ctgcatcaca	gagaaaggaa	gggatccctc	aggagacact	3000
gctgtctcct	tctgggttgt	gctaaacaac	atagggagga	aagctggacc	tggagtcaaa	3060
ggaattgagt	tagtgtgctg	gctctgccat	acttacggca	cccttgggca	ggatatacaa	3120
aggttctcta	cttataaaaat	gggacagtct	aaaactacct	tttagtagag	aagtcaaattg	3180
agaaggtatg	tgaaaactct	gtcaactaaa	tataaagact	aataatttgg	gtattaagag	3240
gctagtttga	gaagccacct	gaattacaca	aacacagcta	cagacatcat	tctgtctaga	3300
gaaagataag	agagaacagg	ttggttgaac	ttgggcagaa	tcacagatac	aattccacac	3360
taaagaatga	aaataagcaa	tgaactagac	agaaggaaga	aatcatgaag	acttaggaag	3420
cagaattaca	atctgtcata	ttaacaaatg	gagtttgcct	tctaagatca	gatgttgctc	3480
agaaactttc	attgtttacc	taataattta	atatcactag	tttcttagtg	ggtcaagcag	3540
atgcaaaatc	cagcttattt	tcttctatgt	gctctcaagc	ttattgctta	ttttaaaagta	3600
aaatcctgaa	aaaggaaaaat	attaggttgg	tgcaaacgta	attgcggttt	ttgcattggt	3660
gaaatttgcc	gttttatatt	ggagtacatt	cttaaataaa	tgtggttatg	ttatacaaaa	3720
aaaaaaaaaa	aaaactcgag					3740

&lt;210&gt; 15

&lt;211&gt; 1196

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 15

ccacgcgtcc	ggtgaatctg	gatgatctat	ttatgtcatt	tatctatgtg	tctcttctta	60
tgccagtacc	ccactgtctt	gattattgtg	gctttatggt	gtcttgaaat	taagtattgt	120
taagctttgc	aatcttttcc	aagattgttt	tggctatagg	tcctttcttt	ataaagttta	180
tattaagctt	ctccattttt	ctttaagaaa	aaatctgctg	gaatggcact	gaatttttag	240
gtcgatttgg	gaagaattga	caatattgaa	cctttcaatc	gatggacatg	gtatgtttct	300
gcattttatt	gggcctttta	aaatttatct	cagcaatatt	ttgtagtttt	agtgaagaag	360
tcttgtatat	atcttttggt	aaatgtattc	ctaaatattt	tgtggaaatg	ttactgtaaa	420
tagtacttta	atttcatttt	taagtattat	gctgggtatac	agaaatatgt	tttatttttg	480
taatatggac	ttttatatte	tgtaatctta	tttaaattca	cttagttcag	tagtagttct	540
tatttttatt	ttattgcaag	tgcccttaggg	ataccttggt	tacacaatcc	atgtcatata	600
tgaataacca	acaggtttac	cttttgtttc	ttcttgcttt	gtcgtaatgg	ctaggacctt	660
ccaacctaaa	tggttgaata	gaagcgggtg	gtgagtttat	tactcttgct	ttattctcaa	720
cagcagagga	tagcgtctag	tttttcta	ttaaaatag	ttagctatag	gttttttccg	780
tttcatgttt	ttatcagatt	gaagatccct	tttattccta	ttttgccaag	agtgtcatga	840
atgcatgttg	aatttcacca	agtacttttt	cctgcactctg	ttgagagaat	cttgactttt	900
ctcttttatt	ttgttaattg	agattcttta	atgttaaact	gatcttgcat	tcctgatata	960
aaccatactt	agtcatgata	tgttatcccc	tttatataat	gctggattcg	gtttgcaaat	1020
attttgtttg	atttttgcac	ctatgctcat	gagacagatt	ggtctgtgat	tttcttttct	1080
tgtaagtctt	gtcaggtgtc	agggccttgg	taattttgac	acattttacaa	acatatattt	1140
ttaagataaa	aataaacttc	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaa	1196

<210> 16  
 <211> 2209  
 <212> DNA  
 <213> Homo sapiens

<400> 16  
 gagctgcgcc gggcttgggc gcctggggcc gccgctcccc accgtcgttt tccccaccga 60  
 ggccgaggcg tcccgagtc atggccggcc tgaactgcgg ggtctctatc gcactgctag 120  
 gggttctgct gctgggtgcg gcgcgcctgc cgcgcggggc agaagctttt gagattgctc 180  
 tgccacgaga aagcaacatt acagttctca taaagctggg gaccccgact ctgctggcaa 240  
 aaccctgtta catcgtcatt tctaaaagac atataacat gttgtccatc aagtctggag 300  
 aaagaatagt ctttaccttt agctgccaga gtccctgagaa tcaactttgtc atagagatcc 360  
 agaaaaatat tgactgtatg tcaggcccat gtcccttttg ggaggttcag cttcagccct 420  
 cgacatcggt gttgcctacc ctcaacagaa cttcatctg ggatgtcaaa gtcataaga 480  
 gcacgggttt agagctgcag ttttccatcc ctgcctgag gcagatcggt ccgggtgaga 540  
 gctgcccaga cggagtcact cactccatca gcggccgaat cgatgccacc gtggctcagga 600  
 ttggaacctt ctgcagcaat ggcactgtgt cccggatcaa gatgcawgaa ggagtgaana 660  
 tggccttaca cctcccatgg ttccacccca gaaatgtctc cggcttcagc attgcaaacc 720  
 gctcatctat aaaacgtctg tgcacatcg agtctgtgtt tgagggtgaa ggctcagcaa 780  
 ccctgatgtc tgccaactac ccagaaggct tccctgagga tgagctcatg acgtggcagt 840  
 ttgtcgttcc tgcacacctg cgggccagcg tctccttcc caacttcaac ctctccaact 900  
 gtragaggaa ggaggagcgg gttgaatact acatcccggt ctcaccacc aaccccgagg 960  
 tgttcaaagt ggaggacaag cagcctggga acatggcggt gaacttcaac ctctctctgc 1020  
 aaggctgtga ccaagatgcc caaagtccag ggatcctccg gctgcagttc caagtttttg 1080  
 tccaacatcc acaaaatgaa agcaataaaa tctacgtggt tgacttgagt aatgagcgag 1140  
 ccattgcact caccatcgag ccacggcccc tcaaacagag ccgcaagttt gtccctgggt 1200  
 gtttcgtgtg tctagaatct cggacctgca gtacgaacct caccctgaca tctggctcca 1260  
 aacacaaaat ctcttctctt tgtgatgatc tgacacgtct gtggatgaat gtggaaaaam 1320  
 ccataagytg cacagaccac cggtagtgc aaaggaaatc ctactcacty cagggtgcca 1380  
 gtgacatcct ycamctgct gtggagctgc atgacttctc ctggaagctg ctgggtgcca 1440  
 aggacaggct cagcctggtg ctgggtgccag cccagaagct gcagcagcat acacacgaga 1500  
 agccctgcaa caccagcttc agctacctcg tggccagtg caccaccagc caggacctgt 1560  
 acttcggctc cttctgccc ggaggctcta tcaagcagat ccaggatgaag cagaacatct 1620  
 cggtgacctc tgcacacctt gccccagct tccgacaaga ggcctccagg cagggtctga 1680  
 cgggtgtcct tataccttat ttcaaagagg aaggcgtttt caccgtgacc cctgacacaa 1740  
 aaagcaaggt ctacctgagg acccccaact gggaccgggg cctgccatcc ctacctctg 1800  
 tgtcctggaa catcagcgtg cccagagacc aggtggcctg cctgactttc ttttaaggagc 1860  
 ggagcggcgt ggtctgccag acagggcgcg cattcatgat catccaggag cagcggaccc 1920  
 gggctgagga gatcttcagc ctggacgagg atgtgtctcc caagccaagc ttccaccatc 1980  
 acagcttctg ggtcaacatc tytaaytgma gccccacgag cggcaagcag ctagacctgc 2040  
 tcttctcggt gatcattacc ccaaggactg tggacttgac tgtcatcctc atgcagcgg 2100  
 tgggaggtgg agtcttactg ctgtctgccc tcgggtctat catttgctgt gtgaaaaaaa 2160  
 aaaaaarama aacaaggggc cccgctgtgg gtatctacaa tggcaacat 2209

<210> 17  
 <211> 1774  
 <212> DNA  
 <213> Homo sapiens

<400> 17  
 taggcacccc aggtttttac actttatggt ccggttcgta tgttgtgtga attgtaaccg 60  
 gataccaatt tcacacaagg aaccagctta tgaccatgat tacgccaagc tcgaaattaa 120  
 ccctcactaa agggaaacaa agctggagct ccaccgcgtt ggcggccgct ctagaactag 180  
 tggatccccc gggctgcagg aattcggcac gacggctgc gggcgcgagg tgaggggccc 240  
 gaggttccca gcaggatgcc cgggtctctg aggaagctga agtgagagg ccggagaggg 300  
 cccagccccg cgggggcagg atgaccaagg cccggctgtt cgggtctgtg ctgggtgctg 360  
 ggtcgggtgt catgatcctg ctgatcatcg tgtactggga cagcgcaggc gccgcgcact 420

tctacttgca	cacgtccttc	tctaggccgc	acacggggcc	gccgctgccc	acgcccgggc	480
cggacagga	cagggagctc	acggccgact	ccgatgtcga	cgagtttctg	gacaagtctc	540
tcagtgtg	cgtgaagcag	agtgaacctc	ccagaaagga	gacggagcag	ccgcctgcgc	600
cggggagcat	ggaggagaac	gtgagaggct	acgactgggc	cccgcgcgac	gcccggcgca	660
gcccagacca	gggcccggcag	caggcggagc	ggaggagcgt	gctgccccgg	ttctgcgcca	720
actccagcct	ggccttcccc	accaaggagc	gcgcattcga	cgacatcccc	aactcggagc	780
tgagccacct	gatcgtggac	gaccggcacg	gggccatcta	ctgctacgtg	cccaaggtgg	840
cctgcaccaa	ctggaagcgc	gtgatgatcg	tgctgagcgg	aagcctgctg	caccgcggtg	900
cgccctaccg	cgaccgcgtg	cgcatcccgc	gcgagcacgt	gcacaacgcc	agcgcgcacc	960
tgaccttcaa	caagtctctg	cgccgctacg	ggaagctctc	ccgccacctc	atgaaggtca	1020
agctcaagaa	gtacaccaag	ttcctcttcg	tgcgcgaccc	cttcgtgcgc	ctgatctccg	1080
ccttcgcgag	caagtctcag	ctggagaacg	aggagttcta	ccgcaagttc	gccgtgcccc	1140
tgctgcggtc	gtacgccaac	cacaccagcc	tgcccgcctc	ggcgcgcgag	gccttcgcgc	1200
ctggccctcaa	ggtgtccttc	gccaacttca	tccagtacct	gctggaccgc	cacacggaga	1260
agctggcgcc	cttcaacgag	cactggcgcc	agggtgtacc	cctctgccac	ccgtgcccaga	1320
tcgactacga	cttcgtgggg	aagctggaga	ctctggacga	ggacgcgcgc	cagctgctgc	1380
agctactcca	ggtggaccgc	cagctccgct	tccccccgag	ctaccggaac	aggaccgcca	1440
gcagctggga	ggaggactgg	ttcgccaaga	tccccctggc	ctggaggcag	cagctgtata	1500
aactctacga	ggccgacttt	gttctcttcg	gctaccccaa	gcccgaaaac	ctcctccgag	1560
actgaaagct	ttcgcggttg	tttttctcgc	gtgcctggaa	cctgacgcac	gcgcactcca	1620
gtttttttat	gacctacgat	tttgcaatct	gggcttcttg	ttcactccac	tgccctctatc	1680
cattgagtac	tgtatcgata	ttgtttttta	agattaatat	atttcaggta	tttaatacga	1740
aaaaaaaaaa	aaaaaaactc	gagggggggc	ccgg			1774

&lt;210&gt; 18

&lt;211&gt; 1674

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (1649)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (1663)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (1665)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 18

caagttggta	cgccctgcagg	taccgggtccg	gaattcccgc	gtcgacccac	gcgtccggtc	60
gaagatagg	tcgagacgg	ggatgttgca	gctgatcatg	cagttgggtt	cggtgctgct	120
cacacgctgc	cccttttggg	gctgcttcag	ccagctcatg	ctgtacgctg	agagggctga	180
ggcacgcggg	aagcccgaca	tcccagtgcc	ttacctgtat	ttcgacatgg	gggcagccgt	240
gctgtgctgc	agtttcatgt	cctttggcgt	gaagcggcgc	tggttcgcgc	tggtggccgc	300
actccaattg	gccattagca	cctacgcgcg	ctacatcggg	ggctacgtcc	actacgggga	360
ctggctgaag	gtccgtatgt	actcgcgcac	agttgccatc	atcggcgggt	ttcttgtgtt	420
ggccagcgg	gctggggagc	tgtaccgccg	gaaacctcgc	agccgctccc	tgacgtccac	480
cggccagggt	ttcctgggta	tctacctcat	ctgtgtggcc	tactcactgc	agcacagcaa	540
ggaggaccgg	ctggcgatc	tgaaccatct	cccaggagg	gagctgatga	tccagctgtt	600
cttcgtgctg	tatggcatcc	tggtccctgg	ctttctgtca	ggctactacg	tgacctcgc	660
tgcccagatc	ctggctgtac	tgctgcccc	tgctcatgctg	ctcattgatg	gcaatgttgc	720
ttactggcac	aacacgcggc	gtgttgagtt	ctggaaccag	atgaagctcc	ttggagagag	780

tgtgggcatac	ttcggaactg	ctgtcatcct	ggccactgat	ggctgagttt	tatggcaaga	840
ggctgagatg	ggcacaggga	gccactgagg	gtcaccctgc	cttcctcctt	gctggcccag	900
ctgctgttta	tttatgcttt	ttgggtctgtt	tgtttgatct	tttgcttttt	taaaattggt	960
ttttgcagtt	aagaggcagc	tcattttgtcc	aaattttctgg	gctcagcgct	tgggagggca	1020
ggagccctgg	cactaatgct	gtacaggttt	ttttcctgtt	aggagagctg	aggccagctg	1080
cccactgagt	ctcctgtccc	tgagaaggga	gtatggcagg	gctgggatgc	ggctactgag	1140
agtgggagag	tgggagacag	aggaagggaag	atggagattg	gaagtgagca	aatgtgaaaa	1200
attcctcttt	gaacctggca	gatgcagcta	ggctctgcag	tgctgttttg	agactgtgag	1260
agggagtgtg	tgtgttgaca	catgtggatc	aggcccagga	agggcacagg	ggctgagcac	1320
tacagaagtc	acatgggttc	tcagggtatg	ccaggggcag	aaacagtacc	ggctctctgt	1380
cactcacctt	gagagtagag	cagaccctgt	tctgctctgg	gctgtgaagg	ggtggagcag	1440
gcagtggcca	gctttgccct	tctgtctgtc	tctgtttcta	gctccatggt	tggcctggtg	1500
ggggtggagt	tccttcccaa	acaccagacc	acacagtcct	ccaaaaataa	acattttata	1560
tagamaaaaa	aaaaaaaaaa	aaggggcgcc	gctctagagg	atccctcgag	gggcccgaagc	1620
ttacgcgtgc	atgcgacgtc	atagctctnt	ccctatagaa	gtngnaaagg	gttc	1674

&lt;210&gt; 19

&lt;211&gt; 2018

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (2010)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (2012)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (2014)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (2016)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 19

cagcaatgaa	atcctgcttt	cttttccctca	gaactactat	attcagtggc	taaatggctc	60
cctgattcat	ggtttgtgga	atcttgcttc	ccttttttcc	aacctttgyt	tatttgtatt	120
gatgcccttt	gcctttttct	ttctggaatc	agaaggcttt	gctggcctga	aaaaggggaat	180
ccgagcccg	attttagaga	ctttgggtcat	gcttcttctt	cttgcggttac	tcattcttgg	240
gatagtgtgg	gtagcttcag	cactcattga	caacgatgcc	gcaagcatgg	aatctttata	300
tgatctctgg	gagttctatc	taccctattt	atattcctgt	atatcattga	tgggatgttt	360
gttacttctc	ttgtgtacac	cagttggcct	ttctcgtatg	ttcacagtga	tgggtcactt	420
gctagtgaag	ccaacaattc	ttgaagacct	ggatgaacaa	atttatatca	ttaccttaga	480
ggaagaagca	ctccagagac	gactaaatgg	gctgtcttca	tcggtggaat	acaacataat	540
ggagttggaa	caagaacttg	aaaatgtaaa	gactcttaag	acaaaattag	agaggcgaaa	600
aaaggcttca	gcattgggaa	gaaatttggt	gtatccccgt	gttatggttc	tccttcttat	660
tgagacatcc	atctcgggtc	tcttggtggc	ttgtaatat	ctttgcctat	tgggtgatga	720
aacagcaatg	ccaaaaggaa	caaggggggc	tggaaatagga	aatgcctctc	ttctacgtt	780
tgggttttgtg	ggagctgcgc	ttgaaatcat	tttgattttc	tatcttatgg	tgctctctgt	840
tgctcggttc	tatagccttc	gatttttttg	aaactttact	cccaagaaag	atgacacaac	900
tatgacaaag	atcattggaa	attgtgtgtc	catcttggtt	ttgagctctg	ctctgcctgt	960

gatgtcgaga	acactgggaa	tcactagatt	tgatctactt	ggcgactttg	gaagggttaa	1020
ttggctggga	aattttctata	ttgtattatc	ctacaatttg	cttttttgcta	ttgtgacaac	1080
atttgtgtctg	gtccgaaaat	tcacctctgc	agttcgagaa	gaacttttca	aggccctagg	1140
gcttcataaa	cttcacttac	caaatacttc	aagggtattca	gaaacagcca	agccttctgt	1200
aaatgggcat	cagaaagcac	tgtgagacgc	acagacggcg	tcttctgcca	ccaagagacc	1260
cgagaactcc	agattcacga	cattcctgtc	ccatgtagaa	gcatttccat	tcaaccgtgg	1320
ccccctttca	gaacctagac	ctatcagtgc	catttttttt	tcataatcta	cgaagaactt	1380
ggctatggct	gatctttttt	aaattttaact	ttctgatgga	ccctgtagtt	tccagtttaag	1440
tgcagattcc	ttacagacat	atagaacagc	gcattcttct	gtagacattt	gctcatgttg	1500
gtaaatacaa	tcacccatat	gaaaaaattg	ttttcacctg	atatgaaaat	gttagaaaag	1560
gcaaactccg	ggactttctaa	agattttactt	aaatcccatt	atgtacttta	ttcagaatgt	1620
agaagctgac	ttgaaaggca	tccttggtac	taagtgaagc	ttattcagaa	aatgcatttt	1680
tcaaattgcaa	tggcaactgc	ttgtagatat	cattttttgca	gtgtatgttg	gagctgtaat	1740
ggttgcaatt	atgtttctta	tttccctaaa	agcaaaaagc	gtagtctctg	atttatgtta	1800
tagaatgata	ctgatttagac	tttgagccaa	ggggaaaata	ctaaattctt	ttaaacctgg	1860
agccttagag	agccacagga	atatcttctg	ttgtacagtc	taataagctg	tggttaggaag	1920
tatcatgtaa	tcacagttta	atgacagttt	atgtatatat	ataattcagt	attccctcga	1980
ggggggggccc	ggtacccaat	tcgcccgaagn	tnancnag			2018

&lt;210&gt; 20

&lt;211&gt; 2098

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 20

ggcacgaggg	accgagctat	tctcctggga	ctggctatga	tggtgtgctc	catcatgatg	60
tattttctgc	tgggaatcac	actcctgcgc	tcatacatgc	agagcgtgtg	gaccgaagag	120
tctcaatgca	ccttgctgaa	tgcgtccatc	acggaaacat	ttaattgctc	cttcagctgt	180
ggctccagact	gctggaaact	ttctcagtac	ccctgcctcc	aggtgtacgt	taacctgact	240
tcttccggggg	aaaagctcct	cctctaccac	acagaagaga	caataaaaaat	caatcagaag	300
tgtccttata	tacctaaatg	tggaaaaaat	tttgaagaat	ccatgtccct	ggtgaatggt	360
gtcatggaaa	acttcaggaa	gtatcaacac	ttctcctgct	attctgacct	agaaggaaac	420
cagaagagtg	ttatcctaac	aaaactctac	agttccaacg	tgctgttcca	ttcactcttc	480
tggccaacct	gtatgatggc	tgggggtgtg	gcaattgttg	ccatggtgaa	acttacacag	540
tacctctccc	tactatgtga	gaggatccaa	cggatcaata	gataaatgca	aaaatggata	600
aaataatttt	tgttaaagct	caaatactgt	tttctttcat	tcttcaccaa	agaaccttaa	660
gtttgtaacg	tgcagctctgt	tatgagttcc	ctaataatatt	cttatatgta	gagcaataat	720
gcaaaaagctg	ttctatatgc	aaacatgatg	tctttattat	tcaggagaat	aaataactgt	780
tttgtgttgg	ttgggtggtt	tcataatctt	atttctgtac	tggaaactagt	actttcttct	840
ctcattccgc	caaaacaggg	ctcagttatt	catttgccaa	gcttcgtgga	ggaatgtagg	900
tgacatcaat	gtgataaagt	ctgtgttctg	agttgtgcga	tctcttgaag	acaatatattt	960
tcatcactta	ttgtttacta	aagctacagc	caaaaatatt	tttttttctt	attctaaact	1020
gagccctata	gcaagtgaag	ggaccagatt	tcctaattaa	aggaagttag	gtacttttct	1080
tgtatttttt	accatatcac	tgtaaagaag	aggggaaacc	cagccagcta	ctttttttca	1140
tcacttttta	ttcataactt	cagattttgt	aaactaattt	ccaaaatata	agctgttttc	1200
attagccagt	tctataatat	cttctgtgta	tttatgtaga	aaatgaacac	accccttttc	1260
catttaagac	cctgctactg	tgtgaagaga	tgatacttac	aaggagtgtc	attacctgtg	1320
agctgactga	atgtttgtag	gtgtccatt	acaatccagg	aaagtctgtg	ttactgatat	1380
ttgtgtggaa	atcttttatt	cacttcaatt	taaccattag	atggtaaaat	taagatgcta	1440
cttgttggtg	aaaattgggtg	gactggtttc	aatgggtaaa	tgtgttgtgg	caaattaatg	1500
tgttgggaata	ttgtcttttg	tgaatttgtg	cttaagtcaa	tgaatgtgta	gtatctcctt	1560
ctgacaagca	ttccctattg	ggatttttaa	gctatgtgca	cagaatatta	gtctcttcta	1620
catgttttat	ttttctattt	ataattccct	ttttgttgt	tatattttat	acacagaata	1680
gatctttttt	ctaaccata	tttgaactga	ataacagact	taaagaaagc	ctttgttcac	1740
attgctattt	acttttgtgt	ttgggggaaa	atacagagga	ttgattttta	ataaaaaaca	1800
ttccatcttt	catttaatat	caatatcaaa	agaagaagac	aaacatctat	ctttctcatc	1860
tatattttaag	tacctttttg	taatgtagta	tcaaagtttt	ttaggtaatg	caaaatttta	1920
caaatcattt	gtggaatgaa	tggtaaaact	aatctgatga	aatggaaaat	tattctgcaa	1980



tattgtaatt catagtttga cttttcataa gcaaataaat ccctaggatg taatcaggac 2040  
 ttcaaatgtg taattaaatt tttttaaaaa aaatctaaaa aaaaaaaaaa aaaaaaaaaa 2098

<210> 21  
 <211> 1746  
 <212> DNA  
 <213> Homo sapiens

<400> 21  
 gcgttcgcgc acctccagct cgggcccgatg tgggaagcttt ggagagctga agagggcgcg 60  
 gcggcgctcg gcggcgcgct ctctctgctg ctcttcgcgc taggggtccg ccagctgctg 120  
 aagcagagggc ggccgatggg cttccccccg ggccgcgcgg ggctgccatt tatcggaac 180  
 atctattccc tggcagcctc atccgagctt ccccatgtct acatgagaaa gcagagccag 240  
 gtgtacggag aggtacagcc ccgacggggc ccgggcaggg agggccgcca ggctggcccg 300  
 ggctggccag ggccttcctg gttggactta tggccgcccc tgggcccgaact agtcggggacc 360  
 tctccgtgtg ccggctgccc tttgagggac acccgcttcc cgggtctgga agggagaagt 420  
 cctcgacgcc gtgccccctt gcagggggag ccccgccccct gccggtgacc cactccgggc 480  
 cgaggctccg aggcgatcca gtccctgattt tcccgtacc gctcgagctc ttgctcctgc 540  
 gcctgcgcgc tttggctcgc cagccgcgcc gccacttcag gtccaggggtg gacgcatgcc 600  
 ctcaggtgcg ggcgtcttgc gagtgcgcct cgcagctctg tggaaagctgc acgcggcttg 660  
 tcggaaaatc aaggcgttct gaggttctaga tgggttaatag cagggtcttc ggtgtctgca 720  
 gtcgacgaac gactggtgta ggcgtttgct gtgagaatgg agaatgcagg ggaacgcccc 780  
 tgactgagaa gcgggcccctg ggaaacgatt gtgaacgcgt gaatgaattg atgactaaaa 840  
 tccgctgcgg gggctcctaca gcgcagatgg taatgccgtt ctgactgggt gggaacggca 900  
 ccttagcaga tacttaaaag gcgccttctg tgtgccactg tcaactgcca cttggtgact 960  
 catttaaaac tcataaccag ccggtgaggt cgggtacttcg ctctcctca ttctgcggag 1020  
 gggaaagcag cacggaaatg ccctgtgact ggcagcggaa aaggcgacca ccgcttgtgt 1080  
 gtgggtgtcc cgacgtccgg agggggcagg agtttccacg ggtcctggga cagagctcac 1140  
 ctgttttgtt ttgaattaca cttatttata tgcaactaca ggcttgacgc tagcgggtgaa 1200  
 gaaggcagat acagcctttt aaggagttgg cagatgagtg ggagagagaa aactaatctc 1260  
 attatcggcc acaggctgtg gtcagtgttt tgaaggaaaa gtacagggat gtttggaac 1320  
 tgtggtattt caggtttgac cttaaactct tacttaaacc agtttttaca aggattggtc 1380  
 taggtgcccg ggcgcggtgc tcacgcctat aatcccagca ctttgggagg ccgagggcgg 1440  
 cggatcacga aatcaggaga tcgagaccgt cgtggctaac acggtgaaac cccatctcta 1500  
 ctaaaagaat acaaaaaatt ggccgggcgt ggtggcgggc acctgtgggt ccagctattc 1560  
 gggaggctgg ggcaggagag tggcgtgaac ccgggaggcg gagctttcag tgagccgaga 1620  
 tcgcgccact gcactccagc ctgggcaaca gagccagact ccgtctcaaa aaaaaaaaaa 1680  
 aaaaagggcg gccgctctag aggatccaag cttacgtacg cgtgcatgag acgtcaatag 1740  
 ctcttc 1746

<210> 22  
 <211> 2876  
 <212> DNA  
 <213> Homo sapiens

<400> 22  
 ccacgcgtcc ggcggcgagg cggcgggtggt ggctgagtcg gtgggtggcag aggcgaaggc 60  
 gacagctcta ggggttggca ccggcccccga gaggaggatg cgggtccgga tagggctgac 120  
 gctgctgctg tgtgcggtgc tgctgagctt ggcctcggcg tctcggatg aagaaggcag 180  
 ccaggatgaa tccttagatt ccaagactac tttgacatca gatgagtcag taaaggacca 240  
 tactactgca ggcagagtag ttgctgggtca aatatttctt gattcagaag aatctgaatt 300  
 agaatcctct attcaagaag aggaagacag cctcaagagc caagaggggg aaagtgtcac 360  
 agaagatata agctttctag agtctccaaa tccagaaaac aaggactatg aagagccaaa 420  
 gaaagtacgg aaaccaggta gtctggacat tttccttgcg ttttgattta tttaggggac 480  
 aactgaaaat tttaagctaa tgaataaaga ggctgaagaa gactggcttc actgattatt 540  
 acccacaat aatatatgga gtgtagtttg gagagaattt ctagatttta atataccaaa 600  
 gttattcact taacagattt gacccaagtt acataagctg aattccatag atgagattgc 660

atztatcttt	ttattaatat	cctgacgttt	cttttttgaa	tcaagtatgt	aaataccttt	720
ctcactgaga	aattttaaaaa	aataatatgt	cagggttagca	tacaaaacag	aaactcagga	780
agaactgtct	aacatgccac	tatgggttatt	ttcaggattg	tggtaggatt	attactacaa	840
taaccttgct	gatcatccta	ggtaatttga	tttaggtgtt	tatccctcag	ctgacgcttg	900
tgtgtacttc	catcctattc	taactttcac	aatcttgaaa	cctctacttt	tcaattttaa	960
acttgacctt	tttcacaata	gtctttctaaa	tattaatcct	tttaaaaaatt	attaatgtat	1020
gcattctacc	atgtcttgtt	gcccttctgt	cattttattat	ttcaatgatc	actgccttat	1080
ctgtttatta	ttccattact	aaccactctc	aggaactgat	ccaaatttga	acttttttaa	1140
aaaatagaat	tttttttttac	tacaaatata	ccattctcat	gaatagaaaa	ctattttttaa	1200
aaaattaaaa	gtacatgaat	cccatacaact	agaaataaca	gctgctcaca	tcattggagaa	1260
tattctctca	ggattttttt	aggtgtatca	gtagttttaa	gaaaagttaa	ttgggtcaaa	1320
atgtgtattc	tatgttgcag	cctgcatttt	gcacttaaca	gtttaccatg	aatgtttttc	1380
catgtcattt	agcttatttt	caatttgata	gttaattggct	ataaaaaatt	ttattttgtaa	1440
atatatggta	ccataaacca	aaacgtttat	gttttgctgg	gagatcattt	tagatgtatc	1500
tttgtgccta	tctgtgaaaa	tttccttaga	attcttaaag	ttaatttgc	gaattggatg	1560
taaaagttta	aaagacagta	gatacacatt	gttaaattgt	tcttcagaaa	agtgtattctg	1620
tttacttcct	ctgaaaatac	ctctttccct	gtacctcacc	caagtgtgta	ttaccatttt	1680
aaaattttta	ctaaatataa	tttaattcat	tcttcatttt	tttaattagg	attttttttt	1740
ttttttttta	gttttttatat	tttttagagac	atgggtctcac	tctgtcactc	aggctggagt	1800
gcagtggcac	gatcatagtt	cactgcagcc	ttgaactcag	gtgatctccc	tctttggcct	1860
cccaaagtgc	tgggattaca	ggcatgagac	actacatcaa	gcctgggaaa	tttttaaaca	1920
tgcataaaaag	tagagggtaa	aatgacaccc	cttaccaccc	ctcactcagc	ttcaataaac	1980
atcaacattc	tgtctgtctt	ttcatcttta	ctaccctaca	cacatttgcc	ctttttttct	2040
ttcctagaat	attttaaggc	aaattccgca	tgtgctttta	catctctgtg	tctctctcta	2100
gtaagaatgt	tcttttcagc	aaaatcacag	aattagccat	tactccttta	ttacctata	2160
ccagactgt	ttaatttttc	tagttgtctt	aagtgtctgt	ttacaattta	tttgtttgaa	2220
ttcatttcta	agctactcag	gaggctgagt	caggagaatt	gcttgaaccc	aagaggcaga	2280
ggttgacgtg	agctgagatc	gtgccactgt	actccagcct	gggtgacaaa	gtgagactcc	2340
atgtcaaaaac	aaaacaaaac	aaaacaaaac	aaaacaaaac	aaacaacaca	attgaataat	2400
atccctcagt	tattattttag	agaagagatg	actcaatata	aatgtaatat	gtgatcattt	2460
ggtcacagaa	tgaatttcag	tcattgctta	ggtaagcctg	gtaaaaaaa	aaaaaaacgg	2520
gattccatga	ttagtgtgtt	gaaaaggggg	ttgccaaaag	gtaacagaca	gagccgggtg	2580
cagtggctca	cgctgtaat	cccagaactt	tgggaggcca	aggcaggcgg	atcacagggt	2640
caagagattg	agacaatcct	ggccaacatg	gtgaagccct	gtatcaacta	aaaatacaaa	2700
aacaatttagc	tgggctgtgt	gacatgtgcc	tgtagtccca	gctactcagg	aggctgaggc	2760
aggagaattg	cttgaacctg	ggaggcggag	gctgcagtga	gctgagattg	ctccactgca	2820
ctccagcctg	gccacagagt	gagattccat	ctcaaaaaaa	aaaaaaaaaa	aaaaaa	2876

&lt;210&gt; 23

&lt;211&gt; 1052

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 23

tcgacccacg	cgctccgcca	cgcgtccgcc	cacgcgtccg	ggcggggcgca	ggacgtgcac	60
tatggctcgg	ggctcgtcgc	gccgggttgc	gcggtctctc	gtgctggggc	tctggctggc	120
ggtgctgcgc	tccgtggccg	gggagcaagc	gccaggcacc	gccccctgct	cccgcggcag	180
ctcctggagc	gcggacctgg	acaagtgcac	ggactgcagc	acctcctgcc	ccctcccgcc	240
tgctttggcc	catccttggg	ggcgctctga	gcctgacctt	cgtgctgggg	ctgctttctg	300
gcttttttgt	ctggagacga	tgccgcagag	agagaagtcc	accaccccca	tagaggagac	360
cggcgagag	ggctgcccag	ctgtggcgct	gatccagtga	caatgtgccc	cctgccagcc	420
ggggctcgcc	cactcatcat	tcattcatcc	attctagagc	cagtctctgc	ctcccagacg	480
cggcgggagc	caagctctct	caaccacaa	gggggtgggg	ggcgggtgaat	cacctctgag	540
gcctggggcc	agggttcagg	ggaaccttcc	aaggtgtctg	gttgccctgc	ctctggctcc	600
agaacagaaa	gggagcctca	cgctggctca	cacaaaacag	ctgacactga	ctaaggaaact	660
gcagcatttg	cacaggggag	gggggtgccct	ccttctctaga	ggccctgggg	gccaggctga	720
cttggggggc	agacttgaca	ctaggcccca	ctcactcaga	tgtcctgaaa	ttccaccacg	780
ggggctcacc	tgggggggta	gggacctatt	tttaacacta	gggggctggc	ccactaggag	840

ggctggccct	aagatacaga	ccccccaac	tccccaaagc	ggggaggaga	tattttat	900
ggggagagtt	tggaggggag	ggtggggggg	gggaaaaaaa	aataaaaaaa	aaaattttta	960
atttttaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1020
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aa			1052

<210> 24  
 <211> 1541  
 <212> DNA  
 <213> Homo sapiens

<400> 24

ggcacgagct	tccctgatcc	tactctgtct	ctctttgttt	tggagaagtt	caatttacc	60
gctggatatg	tgggactagt	attcctgggt	atggcactgt	cctatgccat	ctcttcacca	120
ctatttgggtc	tcctaagtga	taaaaggcca	cctctaagga	aatggcttct	ggtgtttggc	180
aacttaatca	cagccgggtg	ctacatgctc	ttagggcctg	tccaatctt	gcatattaaa	240
agtcagctct	ggctgctggt	gctgatatta	gttgtaagt	gcctctctgc	tggaatgagt	300
ataattccaa	ctttcccga	aattctcagt	tgtgcacatg	aaaatgggtt	tgaagaggga	360
ttaagtacat	tgggacttgt	atcaggctct	tttagtgcaa	tgtggtcaat	tgggtgctttt	420
atgggaccaa	cgctgggtgg	atttctgtat	gagaaaattg	gttttgaatg	ggcagcagct	480
atacaagggtc	tatgggctct	gataagtggg	ttagccatgg	gcttgtttta	tctactggag	540
tattcaagga	gaaaaagggtc	taaatctcaa	aacatcctca	gcacagagga	ggaacgaact	600
actctcttgc	ctaatagaaac	ctagtccgat	ggatccctgga	ttgatacaag	gttgagaaat	660
gaatgctcct	ggccttaaac	atcacccgtag	gaagggtttt	taaaatttta	cgcgcaaaac	720
tccgtggacc	ccgtgccagt	gtcttggaag	tgtcaacgtg	tttttggatg	atcctgtatt	780
gggctgtact	tactgtgata	ctgaaaagct	gtcctgtctga	agcagctata	tttgaaatat	840
taagtatgaa	aggagtaatt	aaaaacaagc	aaaacaaaac	aagacttagt	ttttaaatga	900
ccaaacttgt	ccttaaagat	gttggttatta	actcgagtta	gttcttattt	cctctgttta	960
tttttttattc	taagtacact	gattctgtga	atgtaccttt	tttattacac	agggaaagaa	1020
atgaattaat	ttgatattgt	ctaaatacat	aaagggtgctt	caaaatatgt	agaaacatta	1080
ctatgaaatc	agttttttaa	agatatactt	tctctttgtc	ctgagggtttt	tcgggtctgt	1140
tcaaaaaggaa	gaattcttgc	ctgccataca	gaaactctct	agcactccct	gaccttaagc	1200
ttttctaaaa	attctgtttg	tgtgaaaagt	acaagaataa	caatacttac	aacttccatt	1260
tttgtaacct	acgttcactt	atgatctgga	tttataaaca	ttacttggtg	taacgttttt	1320
catttccttt	aatgtctctg	ttttttggct	ctaccatctg	ttttgttttt	gttttttct	1380
atatcttggg	agatgtattt	catccctaga	gcagggtcagc	ctccttcccc	taatgcgaat	1440
gcttggtttg	ttagggaagg	gcttcctcca	acttcgtgtg	aaattgtgat	gttgaagtga	1500
ataaatgtct	attgtgtaac	aaaaaaaaaa	aaaaaaaaaa	a		1541

<210> 25  
 <211> 2079  
 <212> DNA  
 <213> Homo sapiens

<400> 25

ggcacgaggg	aggcgcggt	gcgggacctg	actagattct	acgacaaggt	actttctttg	60
catgaggatt	caacaacccc	tgtgggcta	cctctgcttg	catctactct	catcaaagc	120
ctgcagtctg	actggaggaa	tgtggtacat	agtctggagg	ccagtggaga	catccgagct	180
ctgaaggatg	gctatgagaa	ggtggagcaa	gaccttccag	cctttgagga	ccttgaggga	240
gcagcaaggg	ccctgatgag	gctgcaggac	gtgtacatgc	tcaatgtgaa	aggcctggcc	300
cgagggtgtc	ttcagagagt	cactggctct	gccatcactg	acctgtacag	ccccaacgg	360
ctcttttctc	tcacagggga	tgactgcttc	caagttggca	aggtggccta	tgacatgggg	420
gattattacc	atgccattcc	atggctggag	gaggctgtca	gtctcttccg	aggatcttac	480
ggagagtggg	agacagagga	tgaggcaagt	ctagaagatg	ccttggaatc	cttggccttt	540
gcttatttcc	gggcaggaaa	tgtttcgtgt	gccctcagcc	tctctcgga	gtttctctc	600
tacagcccag	ataataagag	gatggccagg	aatgtcttga	aatatgaaag	gctcttggca	660
gagagcccca	accacgtggt	agctgaggct	gtcatccaga	ggcccaatat	acccacctg	720
cagaccagag	acacctacga	ggggctatgt	cagaccctgg	gttcccagcc	cactctctac	780

cagatcccta	gcctctactg	ttcctatgag	accaattcca	acgcctacct	gctgctccag	840
cccatccgga	aggaggtcat	ccacctggag	ccctacattg	ctctctacca	tgacttcgtc	900
agtgactcag	aggctcagaa	aattagagaa	cttgacagaac	catggctaca	gagggtcagt	960
gtggcatcag	gggagaagca	gttacaagt	gagtaccgca	tcagcaaaag	tgcctggctg	1020
aaggacactg	ttgacctaaa	actggtgacc	ctcaaccacc	gcattgctgc	cctcacaggc	1080
cttgatgttc	ggcctcccta	tgacagtat	ctgcagggtg	tgaactatgg	catcggagga	1140
cactatgagc	ctcactttga	ccatgctacg	tcaccaagca	gccccctcta	cagaatgaag	1200
tcaggaaacc	gagttgcaac	atztatgatc	tatctgagct	cgggtggaagc	tggaggagcc	1260
acagccttca	tctatgccaa	cctcagcgtg	cctgtgggtta	ggaatgcagc	actgttttgg	1320
tggaaacctgc	acaggagtgg	tgaaggggac	agtgcacacac	ttcatgctgg	ctgtcctgtc	1380
ctggtgggag	ataagtgggt	ggccaacaag	tggatacatg	agtatggaca	ggaattccgc	1440
agaccctgca	gctccagccc	tgaagactga	actgtttggca	gagagaagct	ggtggagtcc	1500
tgtggctttc	cagagaagcc	aggagccaaa	agctggggta	ggagaggaga	aagcagagca	1560
gcctcctgga	agaaggcctt	gtcagctttg	tctgtgcttc	gcaaatacaga	ggcaaggagg	1620
aggttggttac	caggggacac	tgagaatgta	catttgatct	gccccagcca	cggaaagtcag	1680
agtaggatgc	acagtacaaa	ggagggggga	gtggaggcct	gagagggaag	tttctggagt	1740
tcagatactc	tctgttggga	acaggacatc	tcaacagtct	caggttcgat	cagtgggtct	1800
tttggcactt	tgaaccttga	ccacagggac	caagaagtgg	caatgaggac	acctgcagga	1860
ggggctagcc	tgactcccag	aactttaaga	ctttctcccc	actgccttct	gctgcagccc	1920
aagcagggag	tgtccccctc	ccagaagcat	atcccagatg	agtgggtacat	tatataagga	1980
tttttttttaa	gttgaaaaca	actttctttt	ctttttgtat	gatgggtttt	taacacagtc	2040
attaaaaatg	tttataaatc	aaaaaaaaaa	aaaaaaaaaa			2079

&lt;210&gt; 26

&lt;211&gt; 1947

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 26

tgtaaacaga	ttggagaatc	tagcaataag	attcaaagct	aatctggagc	ataaaggcac	60
agttcagaga	cagaataaca	gggatcacaa	gcatgaatta	aaaggaattt	atttgcttca	120
agttcctaga	tacaaccttc	ccatgctgca	cttctccact	gtcggagcac	gttccgaaaa	180
acagaatgcc	ttgatccctg	gtgggtgcga	aggcagttgt	tagggatggc	aggcattgggt	240
gggctccaaa	agatgaaggc	cccacacaca	ggtgtgctgc	atttgggata	tgtgtgggtg	300
tttcttggac	cctttcttct	gggagtaggg	tacacactaa	cgtttaatcc	gctgtctggg	360
tgcatgtcca	cagtacgggtg	gctaaactcg	aacatcactg	caaataggac	gctgagcagg	420
tccgtctgtc	atgtcacgcc	actgcacagg	tccttgtccc	cacacgacgg	ggagtacttg	480
cgtcagatgt	tattgaatag	ctcgtctcgg	gcaggggaag	cggggagttg	gggatattaa	540
ttgggggttt	tattctctatt	atcatgtcag	ctgacattat	gactatataa	tgtagttaga	600
gacaattttt	atcttgcctt	tagtaaaagt	tcagctgcc	aattgtaaat	cattctaatt	660
tggcaggctt	atttttgaca	ttggaaaggg	cagaaagcga	tttgccccag	tagtgtaata	720
ggagttatag	accagaggct	gaaacccaaa	ctatataaaa	aggaattcag	tggagggggc	780
tttgtaatct	ccattaattt	gtgttgctac	ttccaggatc	accaaaaatt	acatgtaatt	840
ttacatgtta	aacacattga	aacataacct	atgtttataa	agcataacgg	gcttcccttc	900
cagaagctct	cctgcttgct	atgaagttag	aacaatgaaa	agtcatagca	gatactcagt	960
ttaaactctgt	gtagaacctg	gtagtgtttg	agctgttatt	cagatttgaa	ttcagactgt	1020
gtgttggtttg	cttatggaca	ctgcctgtcg	ttctgtcact	gttaaattaa	tgagtctata	1080
aggtttttct	tccagaggcc	ataggtgaca	tcactaaaaat	tgcaagataa	attgtaatct	1140
ttgctgctgc	tgactcccc	aacctctccc	ccaccctccg	tgggtgtgctg	ctttctagat	1200
gagcgtgttt	tggagcaggc	ccatctggga	cactctatgc	tttcaccaag	gaagtgcgat	1260
ctgagcagcc	acaatccagc	caaaagagga	tcgtagatat	ttgctctgat	caactagatg	1320
aaaaatatag	agaatggatt	tagccctactg	ctctgtttta	tccaactgag	tctctgacca	1380
gcaattgggtg	cataattatt	acagcaaaaag	ttaagaaatg	aaactgtagc	aattatgtaa	1440
atgaatgtgt	tggcctctta	atacctgtta	ctagtggact	tctgtgagg	aagttagttt	1500
tttgttttga	tgaatgctt	tcgtttttta	aatcttaatt	ctgctgtcca	catcctccca	1560
aagtgtgctt	acttcatttg	tttaatttaa	atgaactttc	ctccttgtat	gtatgagggtg	1620
acttgggtggg	tgggggtgggt	ggtttttgtt	tttgtgtttt	ttctttctta	gggcatctgt	1680
aggcctcaaa	ggacctttcc	tttaggtcat	attcctcaga	aagtcttcaa	tcttcccttg	1740

tttttgtttg	tttgtttttc	ttaaagaata	ttttcaaagc	ttaaatttgt	atattaattt	1800
aggactat	agaagtata	gctgtcgtt	gcggcagcag	tatattctga	aatgtctcat	1860
agatatatat	ttttgaataa	agatggtgtt	gttgaacaaa	aaaaaaaaaa	aaaaaaaaaac	1920
tcgagggggg	gcccgggtacc	caattc				1947

&lt;210&gt; 27

&lt;211&gt; 3379

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 27

atgatattca	agtcgaattc	ccctcactaa	agggaaaaaa	gctggagctc	caccgcgggtg	60
gcggccgctc	taaacctact	ggatcccccg	ggatgcagga	attcggcacg	aggtctcggc	120
tccctccggc	acccgcccgg	cccggctgct	cccggctcct	cccggccatg	gggagctgcg	180
cgcggctgct	gctgctctgg	ggctgcacgg	tggtggccgc	aggactgagt	ggagtagctg	240
gagtgaagttc	ccgctgtgaa	aaagcctgca	accctcggat	gggaaatttg	gctttggggc	300
gaaaactctg	ggcagacacc	acctgcgggtc	agaatgctac	cgaactgtac	tgcttctaca	360
gtgagaacac	ggactctgact	tgctcggcagc	ccaaatgtga	caagtgtcaat	gctgcctatc	420
ctcacctggc	tcacctgcca	tctgccatgg	cagactcacc	cttcgggttt	cctcgcacat	480
ggtggcagtc	tgccggaggat	gtgcacagag	aaaagatcca	gttagacctg	gaagctgaat	540
tctacttcac	tcacctaat	gtgatgttca	agtccccag	gcccggctgcc	atgggtgctgg	600
accgctccca	ggactttggg	aaaacatgga	agccttataa	gtactttgcg	actaactgct	660
ccgctacatt	tgccctggaa	gatgatgttg	tcaagaaagg	cgctatttgt	acttctaaat	720
actccagtc	ttttccatgc	actggaagga	aggttatatt	caaagctttg	tcaccaccat	780
acgatacaga	gaacctttac	agtgcctaac	ttcaggagca	gctgaagatc	accaacctcc	840
cgcgtgcagc	tgctgaaacg	acagtcttgt	ccctgtcaga	gaaatgacct	gaacgaagag	900
cctcaacatt	ttacacacta	tgccatctat	gatttcattg	tcaagggcag	ctgcttctgc	960
aatggccacg	ctgatcaatg	catacctgtt	catggcttca	gacctgtcaa	ggccccagga	1020
acattccaca	tggtccatgg	gaagtgtatg	tgtagcaca	acacagcagg	cagccactgc	1080
cagcactgtg	ccccgttata	caatgaccgg	ccatgggagg	cagctgatgg	caaaacgggg	1140
gtccccaacg	agtgcagaac	ctgcaagtgt	aatgggcatg	ctgataacctg	tcacttcgac	1200
gttaatgtgt	gggaagcctc	agggaaatcg	agtgggtggtg	tctgtgatga	ctgtcagcac	1260
aacacagaag	gacagtattg	ccagagggtc	aagccaggct	tctatcgtga	cctgcggaga	1320
cccttctcag	ctccagatgc	ttgcaaaccg	tgcttctgcc	atccagtagg	atcagctgtc	1380
cttctctgcca	actcagtgac	cttctgcgac	cccagcaatg	gtgactgccc	ttgcaagcct	1440
ggggtggcag	ggcgacgttg	tgacaggtgc	atgggtgggat	actggggctt	cggagactat	1500
ggctgtcgac	catgtgactg	tgacggggagc	tgtgacccta	tcaccggaga	ctgcatcagc	1560
agccacacag	acatagactg	gtatcatgaa	gttctcgact	tccgtcccgt	gcacaataag	1620
agcgaaccag	cctgggagtg	ggaggatgcg	caggggtttt	ctgcacttct	acactcaggt	1680
aaatgcgaat	gtaagggaaca	gacattagga	aatgccaaag	cattctgtgg	aatgaaatat	1740
tcatatgtgc	taaaaataaa	gattttatca	gctcatgata	aaggtactca	tggttaggtc	1800
aatgtgaaga	ttaaaaaggt	cttaaaatct	accaaactga	agattttccg	agggaagcgc	1860
aacattatat	ccagaatcat	ggacggacag	aggatgcact	tgtccaatcc	tcaatcctgg	1920
tttggaatac	cttgtagcag	gacatgagga	tataagaaca	ggcaaaactaa	ttgtgaatat	1980
gaaaagcttt	gtccagcact	ggaaaccttc	tcttggaaga	aaagtcatgg	atatttttaa	2040
aagagagtgc	aagtagcatt	aagatggata	gcacataatg	gcacttggct	atgtccaaaa	2100
cacaaacttt	agagcaagaa	gacctcagac	aggaaactgg	aatttttttaa	agtgccaaaa	2160
catatagaaa	tggttgaaatg	catgggtctt	tatctaattt	atctcttctg	gacccatggt	2220
taaatacagt	tttatttcat	gaagagaaat	gaaaacccct	acactgatat	ctgtttttcta	2280
tgggactgat	tctgaaatct	ttaactatta	agaatatttt	aatagcagca	tgacatttag	2340
cagtaatcca	ttaagggcag	tacctetaac	aaggacgcct	tccagcttca	gctatgttac	2400
ttacggttta	tgctacttaa	agtaatgaat	gacgttttaa	ggaatcccta	accctactat	2460
cagaaaaggt	gtttgttaaa	gagccttctc	ttgtgtgtta	cgcatgaact	ttggtctgta	2520
ggtgttaaat	ggacctctc	catgtgtata	tagtatttcc	ttgtataaag	cacttttacta	2580
cctaccactt	gtgttgtgaa	cgtttggtga	ctgctgttga	aagaaggaaa	agggtgtgtg	2640
agaaagccta	ctgaagcagc	agcactgcca	ctacatgtgg	acaaaagtga	acatatataaa	2700
gaagttgtgc	tattttaactc	tgaatacttg	gagaaactag	gtgaagatgc	aaccagaaag	2760
gagaatatgt	atgcgtgaag	tctcagcttt	gagctggagg	ctagattcca	agatgacagc	2820

catgatgaaa	ctttttaaaa	aactaaacca	gaagagactt	taaaataaga	gaaagaaatc	2880
ataaatgtag	acatatgctt	ggctaaaggg	gaaatggact	ttaaatttta	aagagctcat	2940
ttgcaatgca	cttgatata	cttcaaaaat	tattgtagac	acagaatttg	ttatatTTTT	3000
gtgcttagta	tttaaaccctg	aacattgaaa	cagtttttct	ccttgctctt	cttaacagta	3060
atagtcatta	tatttacctg	ttttttaaca	caatgtatgt	gatagtcaaa	aaatcacagt	3120
ttttcattat	tattcatctt	ctgtacccac	gcataaccac	tatacatagt	ttcttttgta	3180
cttgaatata	caaaacatga	acacagtgcc	atatgaataa	tttcacatac	agaacctttt	3240
tttctctgaa	gtcctgtgga	cttgcaaata	tatatatata	ttgctttgtt	aattttgtttt	3300
tatatttcat	atatgtaata	aaggaatatg	atctgaaaaa	aaaaaaaaaa	aaaaactcga	3360
ggggggggccc	gtacccaaa					3379

&lt;210&gt; 28

&lt;211&gt; 2006

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 28

cccacgcgctc	cgcggaacg	tggtattttt	tcttaagaaa	aaaaaaggaa	gacgttttta	60
taggaaaaga	aaaacaagaa	gcagtagcag	ctaggttaatt	taaactgtctt	cctttttttt	120
tcttaagaga	aaatggaaca	tttaggttaa	atgtctttta	attttaccac	ttacaacac	180
tacatgcccc	taaaatata	ccagtcagta	ctgtatttta	aaatcccttg	aatgatgat	240
atcaggggta	aaattacttg	tattgtttct	gaagtttgct	cctgaaaact	actgtttgag	300
cactgaaacg	ttacaaatgc	ctaataaggca	tttgagactg	agcaaggcta	cttggtatct	360
catgaaatgc	ctgttgccga	gttattttga	atagaaatat	tttaaagtat	caaaagcaga	420
tcttagttta	agggagtttg	gaaaaggaat	tatattttct	tttttctga	ttctgtactc	480
aacaagtctt	gatggaatta	aaataactctg	ccttattctg	gtgagcctgc	tagctaatat	540
aagtattgga	caggtataaa	tttgtcatct	ttaatatagg	taaaatgaat	taagatatta	600
taggattaaa	cataatttta	tacggttagt	actttatttg	ccgacctaaa	tttatagcgt	660
gtggaaattg	agaaaaatga	agaaacagga	cagatatatg	atgaattaaa	aatatataa	720
ggtcaatttt	ggtctgaaat	ccctgaggtg	tttttaacct	gctacactaa	tttgtagact	780
aattttattt	tttagtctag	aaatagtaaa	ttgtttgcaa	gtcactaata	atcattagat	840
aaattatttt	cttgccata	gccgataatt	ttgtaatcag	tactaagtgt	atacgtattt	900
ttgccacttt	ttcctcagat	gattaaagta	agtcacacgc	ttattttagg	aaactggaaa	960
agtaataggg	aaagagattt	cactatttgc	ttcatcagtg	gtaggggggc	ggtgactgca	1020
actgtgttag	cagaaattca	cagagaatgg	ggattttaagg	ttagcagaga	aacttgga	1080
gttctgtgtt	aggatcttgc	tggcagaatt	aactttttgc	aaaagtttta	tacacagata	1140
tttgatttaa	atttgagacc	atagtcagaa	gactcagatc	ataattggct	tatttttcta	1200
tttccgtaac	tattgtaatt	tccacttttg	taataatttt	gatttaaaat	ataaatttat	1260
ttattttatt	ttttaatagt	caaaaatctt	tgctgttgta	gtctgcaacc	tctaaaatga	1320
ttgtgttgct	tttaggattg	atcagaagaa	acactccaaa	aattgagatg	aaatgttggt	1380
gcagccagtt	ataagtaata	tagttaacaa	gcaaaaaaag	tgctgccacc	ttttatgatg	1440
attttctaaa	tggagaaaca	tttggttgca	tccacataga	cctttatgtt	ttgttttcag	1500
ttgaaaactt	gcctcctttg	gcaacattcg	taaatgaagc	agaatttttt	tttctctttt	1560
ttccaaatat	gttagttttg	ttcttgtaag	atgtatcatg	ggtattgggtg	ctgtgtaatg	1620
aacaacgaat	tttaatttagc	atgtggttca	gaatatacaa	tgtaggtttt	ttaaaaagta	1680
tcttgatggt	tcttttctat	ttataatttc	agactttcat	aaagtgtacc	aagaatttca	1740
taaatttggt	ttcagtgaac	tgttttttgc	tatggtaggt	cattaaacac	agcacttact	1800
cttaaaaatg	aaaattttctg	atcatctagg	atattgacac	atttcaattt	gcagtgtctt	1860
tttgactgga	tatattaacg	ttcctctgaa	tggcattgat	agatgggttca	gaagagaaac	1920
tcaatgaaat	aaagagaata	tttattcaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1980
aaagggcggc	cgctctagag	gatccc				2006

&lt;210&gt; 29

&lt;211&gt; 3070

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 29

gttggagcct	tagaccttcg	aagtgcaccta	tagaaggtag	gcctgcaggt	accgggtccgg	60
aattccccggg	tcgacccacg	cgcccgcaac	gtatgaggag	tctgatttct	tcacatcctt	120
gtcaacattt	gttgttgttg	ttgttgttgt	tgtttttaat	tttagctatc	ctagtagatg	180
tgaagtggta	tctcgtgctt	tttatttgca	tttctcta	gactagtgat	ggtgaacatc	240
ttttcatgtg	cttgttgcc	attcgtatat	cttcttgag	aaatgtctat	taagtcctt	300
gccccatttc	aaattggctt	gtctttttgt	tgagttataa	ccattgaatt	ttaattgctg	360
taaaacactc	catgagtaca	tgtagcatgc	actcaataac	tattagctaa	tactatagtt	420
gtaattattt	ttgccacatg	tggcctgaat	tggaaggaaa	ttgcctagtt	tttccaagta	480
ggagtggaga	tgtaacaaaa	cctgtattta	aggccaataa	agatttgact	ctgtcagttg	540
tgatactttg	cttttaattg	atgctcccaa	atcaagaaat	tagaaacgtg	tatacataag	600
aatcagttta	tctttggcag	aaatcagtat	attatcaatg	tttaattaat	tcctaagaat	660
taatggcaat	agtgcctcta	gttttagctg	agatttagat	catctttctt	cattctgccc	720
tgttatgcga	catttgctat	aaatgccaca	atgagaattt	tttgaaacct	gggtcagaga	780
agcctacttt	ttatatattaa	taagtatgtg	attgatgtat	tatcattggg	ttatgcttat	840
tgcactcttg	ttttatgggg	tgttcaaaa	aagatagatg	agaatagggtg	atccaaagta	900
atttagaatt	ttatcctcca	gttgggactg	aaccatcttg	aactgatgaa	tatattctac	960
caggtttata	ggtagatatg	cgttttcatt	atatcggtag	tttggagtga	gacagtagct	1020
tttaagcaga	acaacagcca	tatgtaagtc	taaatttggg	tattttcttt	gaataagcaa	1080
tggaaactgt	atgcattcag	gaactgggtga	ttcagatata	ttctgaaggc	cctctctccc	1140
ctgtgctccc	acagcttttg	tcctgaactc	tcctgtctcc	ttgcgatgcc	tctgtcctta	1200
tacctatttt	gtgaactctg	cctgcaatgc	cgccccaggt	cagttgggtga	ctcctcaaac	1260
ctcactgaag	tctttcgggg	tggattcaaa	tgctgctttt	agccattaaa	atctgttaca	1320
tctgtttacat	actttcacaa	tagcacttat	cctagtgtag	ttttaaaaat	gcacttaaaa	1380
atttttaaa	cagtgcattg	atgcagtaaa	caaaattgga	cagtacagaa	ataaatgggtg	1440
aagacagtct	cccagttttc	ttccccatgg	atagcacttt	ctctttttta	tccttccaaa	1500
gatattttgt	attacagctg	gaatttttat	accgtatgaa	gtgcttatct	atcttatctt	1560
ttacttcagg	gtaatttatg	agtaatgaaa	aaattcaagc	attacaaaaa	tgtagagtgt	1620
gtgaagcaaa	aggtccccct	tactttgatc	cctaacacta	gccctttcca	aataacactg	1680
tcctaattaa	cattcctctc	tcagaagta	acctctgtta	taagccttgt	ctgtgccttc	1740
agattgttct	gcttacactt	atatgcctat	ttaaaagtgt	atatttttgt	attttgtttg	1800
tttgtttgtt	tgagacagag	tttcgctctt	gttgcccagg	ctgaagtgca	atggcactat	1860
ctctgctcac	tacaacctgt	gtctccctgg	ttcaagtgat	tctcctgggt	cagcctcctg	1920
agtagctagg	attacaggtg	cacgccacca	tgcccggtat	atttttgtat	ttttagtaga	1980
gatggagtgt	caccatgttg	gccaggctgg	tcacgaactc	ctgacctcag	gttatccacc	2040
caccttggtc	tttcaaaactg	ctgggattac	aaggcatgag	ccactgtgcc	cggccaagaa	2100
ttttttttat	gataccatag	tgagctctct	gcctcttcgg	aacgatgtcc	actttgctta	2160
tgatcaaccc	aagcaggact	cttctctccc	tggacgcctc	ccccctgggtc	tggaatcttc	2220
cagttctgcc	agaattggcc	tttcccagat	gctgcaaaact	tccagttgaa	cccccttttc	2280
tgtgtggccc	ctggggctgc	gagacaaaa	tccatgagtt	ctgtgtacct	tagacctttg	2340
gaaggtgaga	gcagggccct	gagaaaaggc	aaaacgcctc	ctctccctga	tgataccctt	2400
accgcctcta	ctcctcacca	gaattgtcag	tggcctttca	ccacagtggg	ccttccctgct	2460
tgagccctgc	actgtcccag	accacacaga	agtctgggtc	cctctggggc	cctgggatgg	2520
tcaccgaaga	gaagcacgct	gtccccgtct	tctcctgggt	tctgccagaa	aatcgaacaa	2580
gtgcaattaa	cacactgtta	ctgccgaagc	ctgaaactcc	caggacttgt	ccttgatcct	2640
tccagaaacc	accaggtccg	gcacttggag	ccccccggag	agggacctcc	cagccgagcc	2700
ctcaaagaac	tccatgaaat	caggaactgc	ttgatgaaat	gtatctcctt	gtacctggaa	2760
gatgaagccc	aaacacccac	acctctgtct	ccccagggtc	tcgggatgtc	tccagcagcc	2820
cggccacgca	gcttcccagg	tgggtctggg	gaggtggggg	cagggacctat	ctctgtcccc	2880
tccacctca	ctccatccac	ctcggagacc	acctcccccc	agccagatac	ggaataaaaac	2940
tacagacgca	gaaaaaaaaa	aaaaaaaaaa	gggcccgcgc	tctagaggat	ccctcgagggg	3000
gccccagctt	acgcgtgcat	gcgacgtcat	agctctctcc	ctatagttag	tcgtattata	3060
aggagacaaa						3070

&lt;210&gt; 30

&lt;211&gt; 2227

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

<220>  
 <221> SITE  
 <222> (289)  
 <223> n equals a,t,g, or c

<400> 30

cgagcgcgtg	ggcgcaccca	cgcgcccg	aaaaarggaa	aaatgcccgt	gtaaaatctc	60
gttctgtgtc	tgaattgccg	taggctcaga	tcttcatttg	aggttctgtg	tctgaattgc	120
cgtaggctca	gatcttcatt	tgaggttatg	ttctataagt	taacgttgat	cttgtgtgag	180
ctttcggtag	ctggagtaac	acaggcgcc	tcacagcgac	ctctccagcg	ccttccaagg	240
cacatctgca	gccagcgtaa	tcctcctggg	agatgcctcc	tcaaggccnt	gctccagacc	300
acgtggggar	ggcctgacar	ccaattccca	ggctgtcccc	acccttgrag	agtgacccta	360
aacgctagac	agatggggaa	tgggaaagaa	aagaaagctg	cagacctcaa	gttaaaattc	420
cctcaaaaac	gtttttat	atctgctttt	tctgaaagga	taaaggcttt	ttgaaaatta	480
ttttctaaca	aataacatga	acacttctag	aaaccctaga	aaaacacaaa	gtattcaaaa	540
tagaaagaaa	aattacccat	tactctttta	gccagcatta	tccattgcgg	tgcttttgga	600
gttgggtgag	ggcgtagcct	ctgccaaagt	aaggagcccg	gtgggtggctg	tggcattcct	660
gcaggggtgt	tttttttct	ttgagatgga	gtctcactct	tgtaacccca	gctggaatgt	720
ggtggtgtaa	acagctcact	gcagccttga	ccctgagcct	caagcgatcc	ttctgccttg	780
gcctcctgag	tagctgggat	cccaggcgag	agtcaccaca	ccctgtccat	gttccctgcag	840
gtcttgatat	gcgaggacgc	tgtgtcttcc	ctgccacatt	ttcttcttct	ttcttgagac	900
agacccttgc	tccatcaccc	aggccagagt	gtggtsgtgc	gaacacggct	cactgcagcc	960
tcgaccctca	ggctcaagcg	atcctcacgc	ctcgaccccc	caaagtgtctg	ggatcacagg	1020
cgagagtcac	catgttggcc	tgaatcttca	gggtatttta	cggttgaagt	gtcacttact	1080
tarccatssc	tgtttcaaga	gtgtagggtg	tcaccctgtc	tctgycgctg	acctggcctg	1140
gaccctcggc	tgtgagagg	aggggtgggc	tgggctggag	gaacctraag	ccctcgtgat	1200
gtcacaagcc	catctggctg	ggcatccctt	gctgtgtcct	gagctgcaca	tgccccaggt	1260
ggccccccaca	gcagaggcga	gccactgrag	ggtgragggc	ttccacggac	ggtcttcagg	1320
ggragaagaa	gggcccaggc	ccccaggaga	ctcaggagac	cagagcctgg	ggtcaggggc	1380
tyagcagggg	ctyarccagg	gctggatgtc	cggagccagc	cccgmagccc	tgkgktcttt	1440
gttcttcgca	ctcccaccgt	ccgtgtgaac	agctccagcc	ccacctgcgc	ctccctgtgc	1500
tgggtcccat	caggagagcc	agaagacgtg	tgtgcttctg	aaattgggtc	cctacatgcc	1560
tttgtccag	tgcaccttgc	tccttccatt	tactatcgag	atttaaatgc	ctgttttctc	1620
cccagaggtt	gacggatata	ttcagacgtt	acgacacgga	tcaggacggc	tggattcagg	1680
tgtcgtacga	acagtacctg	tccatgggtc	tcagtatcgt	atgacctggg	cctctcgtga	1740
agagcagcac	aacatggaaa	gagccaaaat	gtcacagtgc	ctatctgtga	gggaatggag	1800
cacaggtgca	gttagatgct	gttcttccct	tagattttgt	cagtggggga	cccagctgta	1860
catatgtgga	taagctgatt	aatggttttg	caactgtaat	agtagctgta	tcgtttctaat	1920
gcagacattg	gatttgggtg	ctgtctcatt	gtgccatgag	gtaaatgtaa	tgttttcaggc	1980
attctgtctg	caaaaaaatc	tatcatgtgc	ttttctagat	gtctctggyt	ctatagtgcg	2040
aatgctttta	ttagccaata	ggaattttta	aataacatgg	aacttacaca	aaaggctttt	2100
catgtgcctt	acttttttaa	aaaggagttt	attgtattca	ttggaatatg	tgacgtaagc	2160
aataaaggga	atgttagacy	tgtaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	2220
aaaaaaa						2227

<210> 31  
 <211> 1288  
 <212> DNA  
 <213> Homo sapiens

<400> 31

cccccggtc	gcaggaattc	ggcagagcc	tgacctccc	agcctcatct	ctccctctc	60
tgtctctgcc	ctctgtgtc	cagccaacgt	ggcctgtcac	tgtccacct	gccatactgt	120
cctgactcca	ggcctttgcc	tgtgtctatg	cctctgttgg	gacctctct	gtctctcccc	180
tgtgtgtct	gctaattccc	actcgtgtca	gtgatccatg	gctgcagaac	acaccactcc	240
atccatggaa	aacaatcaca	atcattgatt	actatctctc	cctgggcttc	ctggggtgga	300
ctgggctcag	ctgggtggtt	cactttgggg	cctcagcagt	catgggcaga	cagtggctgg	360



ggtcactgca	aagacttccc	tgcactctctg	gcagttgatg	ctggctgtca	tctgagacac	420
ctacccaggg	cctctccctg	gggcctgggc	tcctgcttag	cttggttggg	aggctccaag	480
accaacatcc	caagaaagat	gagacagaag	ccagatcacc	tttttggggc	tggcttcaga	540
agtcacccag	catcacttct	gctgcattta	tttcttaaaa	cacaaatatc	aaacccccatc	600
tcttgatggg	agggggcctc	atggtttgta	aacatgttct	aaactccact	ctgcccggcc	660
ttggctcaac	gtgtctggta	atgtgtgggc	tgtgaggctc	cccgaacgta	gacctcagac	720
tgcaacgctg	gccgttacag	ggtctggcac	acggggccac	gtcaggccca	tcgccacagt	780
gatggttgtt	ctgtgactgt	ttctgggtggc	ctctgctcca	cactccaggc	tgacgctgtg	840
ccccttccac	tgggacccctc	gggtgggttc	catgcacttg	tgcctaaat	cctgctccta	900
gactaaactt	catctcctgt	gttctcattc	tgcagcatgg	ctgttaggga	acctgaccat	960
ctgcagcgcg	tctcgttgcc	aagggtataat	gtcagtgcct	cccttcagt	gtccccatgt	1020
cacagaattg	tcctgcagcc	ctggcacatg	tgtgccattg	gggagctggg	gcagggtcctc	1080
tttcatcctg	tggctccgag	ggagggggcc	gtcctctccc	cagtctctac	cctgacttgg	1140
ccctcgtcct	gcagccactc	agagagcacg	atggagctgg	agcttcagtt	ttgaccaaatt	1200
gcgtgtcgcc	ggctttttgtg	tgtgtgtgtg	tgtgacagag	ccagaccctg	tcttttaaaaa	1260
aaaaaaaaaa	aaactcgagg	ggggggccc				1288

&lt;210&gt; 32

&lt;211&gt; 3280

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 32

gcgtccgggtg	agaggcaagg	acttttctatc	tttagggcat	tctgaccacg	tcctgcttca	60
gagagattgt	tcccggcgctc	tcagtgcctat	ggggagcagg	ttcctcctgg	tcctgctctc	120
aggtctcact	gtcttacttg	ctctgccagg	atcagaagcc	aagaattctg	gagcttccctg	180
tcctccatgc	cctaaatatg	ccagctgcca	caacagcacc	cactgtactt	gtgaagatgg	240
ctttcggggc	aggtctggca	ggacatactt	tcattgatcc	tctgagaagt	gtgaagatat	300
taatgaatgt	gaaaccgggc	tggcaaaagt	caagtataaa	gcataattgta	ggaataaaagt	360
tggaggttac	atctgtagct	gtttggtaaa	atatacttta	ttcaactttc	tggctgggtat	420
tatagattat	gatacccg	attgttatga	gaacaatagt	caagggacga	cacagtcaaaa	480
cgtggatatt	gggtgagtg	gggtgaagcc	tggatttggg	aaacagctgg	tacgtataac	540
tatgccattt	tcctacccaa	acattaacat	gtcttccctg	gatttttagg	gtagggtagt	600
tctatccagg	ggtaattttg	tcctctgtcc	caaggctatc	tgtcaatgac	tggggacact	660
tttggttgtc	ataccttggg	ggtgatgtgt	gtgactggca	tctggtggat	ggagaccagg	720
gatacagctc	aacatccctac	agtggcccagg	acagcctccc	acaatcaaga	agtgcctagt	780
gccatatgtc	catagagata	gagaaatata	agtgtagggg	gaaagtgcct	cagctggcat	840
tcaaagacct	acatcaagca	cctagattct	caatggcaca	cgcaccttgt	agcaccttaa	900
taaatatcgt	tgccttcttg	gtcctccact	ccaacacttg	tgcataatcc	ctatttctta	960
catttcagta	agagtcaatc	taattagctt	aatttttttg	gaaggaaaat	ctgagaagaa	1020
atggaagcag	agaggacttt	gcaagaaggg	ctactcaact	aattcaaagc	gtggagttga	1080
gcactctggaa	tgcgagtttt	gcttctccag	gaaaggggtca	aatttctgaa	tttgatatag	1140
tctatgaaac	caagagggtgc	aatgagacaa	gggagaatgc	ttttctggaa	gctggaaata	1200
acaccatgga	tatcaactgt	gctgatgctt	taaaaggaaa	cctaagagag	agcactgcag	1260
ttgccctatc	acttatcaat	ctcttgggga	tattctgaat	gcactccttt	ttagtaaaccg	1320
aaaagggatg	caggaagtaa	aactgaactc	ttacgtttgt	agcggcacca	tcggtttgaa	1380
ggaaaaaatt	tcctctctctg	aacctgtgtt	cctgactttt	cgcataatc	agcctggtga	1440
caagagaaca	aaacatatct	gtgtctactg	ggagggatca	gagggaggcc	gctggtccac	1500
ggagggctgc	tctcatgtgc	acagcaacgg	ttcttacacc	aaatgcaagt	gcttccatct	1560
gtccagcttt	gccgtcctcg	tggctcttgc	ccccaggag	gacctgtgc	tgaccgtgat	1620
cacccagggtg	gggtgacca	tctctctgct	gtgcctcttc	ctggccatcc	tcaccttcc	1680
cctgtgccgg	cccatccaga	acaccagcac	ctccctccat	ctagagctct	ccctctgect	1740
cttcttggcc	cacctcctgt	tcctgacggg	catcaacaga	actgagcctg	aggtgtgtgtg	1800
ctccatcatt	gcagggtgc	tgcacttctc	ctacctggct	tgcttcacct	gtagctcct	1860
ggaagggctg	cacctcttcc	tcaccgtcag	gaacctcaag	gtggccaact	acaccagcac	1920
gggcagattc	aagaagaggt	tcatgtaccc	tgtaggctac	gggatcccag	ctgtgattat	1980
tgctgtgtca	gcaatagtgt	gaccccagaa	ttatggaaca	tttactcact	gttggctcaa	2040
gcttgataaa	ggattcatct	ggagcttcat	ggggccagta	gcagtcatta	tcttgataaa	2100

cctggtgttc	tacttccaag	ttctgtggat	tttgagaagc	aaacttttct	ccctcaataa	2160
agaagtttcc	accatttcagg	acaccagagt	catgacattt	aaagccattt	ctcagctatt	2220
tatcctgggc	tgttcttggg	gccttgggtt	ttttatgggt	gaagaagtag	ggaagacgat	2280
tggatcaatc	attgcatact	cattcaccat	catcaacacc	cttcagggag	tgttgctctt	2340
tgtggtacac	tgtctcctta	atcgccaggt	tcgaatggaa	tataaaaagt	ggtttagtgg	2400
gatgcggaaa	gggttagaaa	ctgaaagcac	tgagatgtct	cgctctacta	cccaaaccac	2460
aacggaagaa	gtggggaagt	cctcagaaat	ctttcataaa	ggaggcactg	catcatcatc	2520
tgcagagtca	accaagcaac	cgcagccaca	ggttcattct	gtctctgctg	cttggtctaaa	2580
gatgaactga	cctggcaagt	gccatggcaa	tgaccgggaa	gttacccttc	ctttccgttt	2640
gtctacagcg	cccctgtggt	cacacataga	ttggacaaat	gccactattt	ctagctttcc	2700
tgtgaaaagt	ctaggctcat	tcacctattt	tggcttttta	tgttcataga	aagaacaaga	2760
catttgggag	aattcttaga	tccagagtcc	agtagtgtgg	cacgtgcaat	gaagtgtcgg	2820
aaggatgcat	tttaaagatg	gcgggcggga	gaagtggatt	ttcttcttgc	agctactgcc	2880
accttgccag	aaactcacta	actggcatct	ggattcagct	catagttccc	tttctggcct	2940
ctctgctgta	ttttatgctc	ccaaagatct	tacattaaca	ctccacattc	acataattca	3000
acaattttca	tatggatcag	tattaaagag	ggtgttgcat	tttgcaatac	aaaaatgcat	3060
tatcaggtgc	tggagaggat	gtggagaaat	aggaacactt	ttacactgtt	ggtgggactg	3120
taaactagtt	caaccattgt	ggaagtcaat	gtggcgattc	ctcagggatc	taggactagg	3180
aataccattt	gacacagcta	tcccattact	gggtatatac	cccaagggct	ataaatcatg	3240
ctgctataaa	gacacatgca	cacgtaaaaa	aaaaaaaaaa			3280

&lt;210&gt; 33

&lt;211&gt; 1297

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 33

ccacgcgtcc	ggactgcttt	acggacattg	gatgaagccg	aagcatttag	aatgggtgct	60
ggcacacagt	tgggtgcgtga	tatgggttaag	ctttgtgtcc	ccaccacat	ctcatcttga	120
atgtgacggg	ttccccggct	ccctcctgcc	gccatgtgaa	gaaggctggt	gcttccccct	180
caccttccac	caccatgatt	gccatggatg	ctccccactc	caaagcagcc	ctggacagca	240
ttaacgagct	gcccagaaac	atcctgtcgg	agctgttcac	gcacgtgccc	gcccgcagc	300
tgtctgctgaa	ctgccgcctg	gtctgcagcc	tctggcggga	cctcatcgac	ctcatgacce	360
tctggaaacg	caagtgcctg	cgagagggct	tcataccaa	ggactgggac	cagcccgtgg	420
ccgactggaa	aatcttctat	ttcctacgga	gcctgcatag	gaacctcctg	cgcaaccctg	480
gtgctgaaga	ggatatgttt	gcatggcaaa	ttgatttcaa	tgggtgggac	cgctggaagg	540
tggagagcct	ccctggagcc	cacgggacag	attttcctga	cccaaagtc	aagaagtatt	600
ttgtcacatc	ctacgaaatg	tgcctcaagt	cccagctggt	ggaccttgta	gccgagggct	660
actgggagga	gttactagac	acattccggc	cgacatcgt	ggttaaggac	tggtttgcgt	720
ccagagccga	ctgtggctgc	acctaccaac	tcaaagtcca	gctggcctcg	gctgactact	780
tcgtgttggc	ctccttcgag	ccccacctg	tgaccatcca	acagtggaa	aatgccacat	840
ggacagaggt	ctcctacacc	ttctcagact	acccccgggg	tgtccgctac	atcctcttcc	900
agcatggggg	cagggacacc	cagtactggg	caggctggta	tgggccccga	gtcaccaaca	960
gcagcattgt	cgtcagcccc	aagatgacca	ggaaccaggc	ctcctccgag	gctcagcctg	1020
ggcagaagca	tggacaggag	gaggctgccc	aatcgcccta	ccgagctggt	gtccagattt	1080
tctgacagct	gtccatcctg	tgtctgggtc	agccagaggt	tcctccaggc	aggagctgag	1140
catggggtgg	gcagtgaggt	ccctgtacca	gcgactcctg	ccccggttca	accctaccag	1200
cttgtggtaa	cttactgtca	catagctctg	acgttttgtt	gtaataaatg	ttttcaggcc	1260
gggcaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaa			1297

&lt;210&gt; 34

&lt;211&gt; 2184

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 34

ggcacgaggg	gcccaggagat	gtgctgctgg	ccgctgctcc	tgctgtgggg	gctgctcccc	60
------------	-------------	------------	------------	------------	------------	----

gggacggcgg	cggggggctc	gggccgaacc	tatccgcacc	ggaccctcct	ggactcggag	120
ggcaagtact	ggctgggctg	gagccagcgg	ggcagccaga	tcgccttccg	cctccaggtg	180
cgcactgcag	gctacgtggg	cttcggcttc	tcgcccaccg	gggccatggc	gtccgccgac	240
atcgctgtgg	gcgggggtgg	ccacgggcgg	ccctacctcc	aggattattt	tacaaatgca	300
aatagagagt	tgaaaaaaga	tgctcagcaa	gattaccatc	tagaatatgc	catggaaaat	360
agcacacaca	caataattga	atttaccaga	gagctgcata	catgtgacat	aaatgacaag	420
agtataacgg	atagcactgt	gagagtgtatc	tgggcctacc	accatgaaga	tgcaggagaa	480
gctgggtcca	agtaccatga	ctccaatagg	ggcaccaaga	gtttgcgggt	attgaatcct	540
gagaaaacta	gtgtgctatc	tacagcctta	ccatactttg	atctggtaaa	tcaggacgtc	600
cccatcccaa	acaaagatac	aacatattgg	tgccaaatgt	ttaagattcc	tgtgttccaa	660
gaaaagcatc	atgtaataaa	ggttgagcca	gtgatacaga	gaggccatga	gagtctgggtg	720
caccacatcc	tgtctatcca	gtgcagcaac	aactttaacg	acagcgttct	ggagtccggc	780
cacgagtgtc	atcaccccaa	catgcccgat	gcattcctca	cctgtgaaac	tgtgattttt	840
gcctgggcta	ttggtggaga	gggcttttct	tatccacctc	atgttggtat	atcccttgge	900
actccattag	atccgcatta	tgtgtctcta	gaagtccatt	atgataatcc	cacttatgag	960
gaaggcttaa	tagataattc	tggactgagg	ttattttaca	caatggatat	aaggaaatat	1020
gatgctgggg	tgattgaggc	tggcctctgg	gtgagcctct	tccataccat	ccctccaggg	1080
atgcctgagt	tccagtctga	gggtcactgc	actttggagt	gcctggaaga	gctctggaag	1140
ccgaaaagcc	aagtggaaat	catgtgtttg	ctgttcttct	ccatgctcac	ctggctggca	1200
gagcatcagg	ctgcgtcatt	ttcgaaaagg	gaaggaaatg	aaattacttg	cctatgatga	1260
tgattttgac	ttcaattttc	aggagtttca	gtatctaaag	gaagaacaaa	caatcttacc	1320
aggagataac	ctaattactg	agtgtcgcta	caacacgaaa	gatagagctg	agatgacttg	1380
gggaggacta	agcaccagga	gtgaaatgtg	tctctcatac	cttctttatt	acccaagaat	1440
taatcttact	cgatgtgcaa	gtattccaga	cattatggaa	caacttcagt	tcattgggggt	1500
taaggagatc	tacagaccag	tcacgacctg	gcctttcatt	atcaaaaagtc	ccaagcaata	1560
taaaaacctt	tcttccatgg	atgctatgaa	taagtttaaa	tggactaaaa	aggaaggtct	1620
ctccttcaac	aagctgggtc	tcagcctgcc	agtgaatgtg	agatgttcca	agacagacaa	1680
tgtgtgagtgg	tcgattccaa	ggaatgacag	cattacctcc	agatatagaa	agaccctata	1740
aagccagaac	cctttgggtg	gtggcacgtc	ttcttccctc	ttcccctgcc	acagagattt	1800
ctcccatcca	acttgcttgt	ttgccttctg	ctaactcagc	tgcacgctga	gcaccaagag	1860
cttgtgatca	aaattctgtt	ggacttgaca	atgttttcta	tgatctgaac	ctgtcatttg	1920
aagtacaggt	taaagactgt	gtccactttg	ggcatgaaga	gtgtggagac	ttttcttccc	1980
cattttccct	ccctcctttt	tcctttccat	gttacatgag	agacatcaat	caggttctct	2040
tctctttctt	agaaatatct	gatgttatat	atacatggtc	aataaaaata	aactggcctg	2100
acttaagata	accattttta	aaaattgggc	tgtcatgtgg	gaataaaaaga	attctttctt	2160
tcctaaaaaa	aaaaaaaaaa	aaaa				2184

<210> 35  
 <211> 949  
 <212> DNA  
 <213> Homo sapiens

<400> 35						
ggcacgagct	cccattgacc	ttattgtctca	gagtgacagta	ttagggcaag	gtttcgccac	60
tgcctccctc	catgaatgta	tttctccctc	ctgccctggg	gacatgggga	gtggcccgtt	120
tctttcccca	tctagtccca	gaaagatggt	gtttggtttt	ctgttggttg	attttttttt	180
tttttttttt	tttttgcacc	aaagtggcaa	ctaggtcagt	gttgggggat	caagctggcc	240
tcgggggtgg	gggccccac	ctgcctctcc	ctggttccca	cagtgttagc	gtccctgaaa	300
agacaatatt	ctctctaaag	caataagggg	tgacgggccg	gggggagtg	ttgtgtgtgc	360
tggccccccag	ctcccccttc	ctgccaggtg	tgggggagac	tcctgttggtg	actgaatgta	420
acccccccac	ccctgcgcga	gccaatgcag	gggaaggggg	acactcttcc	tgtctcttct	480
ccccagctaa	agagactttg	gacttagggg	gcccattgagc	ctggagaggc	cttaaccctg	540
tgaggaagta	tagggggagc	cctctccccc	cccatcccc	ttctgagagt	ggtcaatggt	600
tacaagcccc	tgagcccccc	tgccagggga	ctcagacccc	gttgctgtcc	ttccccggcc	660
ccggtcttcc	tgggcccctg	ctgtctctct	gccctttctg	gggttggggg	gggtgcaggg	720
gtcacctgtg	tccctgtctg	ccttgtaccc	acagtctccc	cgccccctct	ccaccctgtg	780
tgacttccct	ctctttttacc	tgtctctgta	aatactccct	tctcccaata	aaacttggtg	840
tgtgttctcc	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	900

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa

949

<210> 36  
<211> 3338  
<212> DNA  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (2861)  
<223> n equals a,t,g, or c

<220>  
<221> SITE  
<222> (3328)  
<223> n equals a,t,g, or c

<220>  
<221> SITE  
<222> (3330)  
<223> n equals a,t,g, or c

&lt;400&gt; 36

taccggggaca	ttcttgaagtc	cgagaagtca	acgggcgcggt	tgctgcgggcc	gccgcgctcc	60
ccggccccgag	gcatgaggaga	acggagcgggt	gtacagcccc	actacggagg	aggacccggg	120
ccccgccaga	ggcccccgga	gcggcctcgc	tgcctacttt	ttcatgggcc	ggctcccatt	180
gctccggcgc	gttctcaagg	gcttgcagct	gttgcgtgct	ctgctggcct	tcctctgtga	240
agaagttgta	tcacaatgta	ctttatgttg	aggactttat	tttttgagt	ttgtaagctg	300
cagtgccttt	cttctgagtc	tccttatact	gattgtgtat	tgactccat	tttatgagag	360
agttgatacc	acaaaagtaa	aatcatcgga	tttttatatt	actttgggaa	caggatgtgt	420
gtttttgttg	gcatccatca	tttttgtttc	cacacatgac	aggacttcag	ctgagattgc	480
tgcaattgtg	tttggaattta	tagcaagttt	tatgttccta	cttgacttta	tcactatgct	540
gtatgaaaaa	cgacaggagt	cccagctgag	aaaacctgaa	aataccacta	gggctgaagc	600
cctcactgag	ccacttaatg	cctaaagact	ctggggagca	gatgttacct	aaggtagtga	660
ccctgcattg	tggtgcctga	gccctggcag	aagctcttgt	aaaatttggt	aattgtttaa	720
accacttctt	ttggagagca	aggggaaggt	caagaaggca	gttttatcaa	tattgtgtca	780
gtcaccacaa	agtaggccag	ataagttaaa	aaaaattttt	ttttaataaa	taattgaaac	840
ttatctcaaa	tggagatttt	ggtgggagga	ggagaaaaca	attgttttta	aatcacacag	900
ctcaacgggt	gataaatgat	tctgtcattc	tgttacaggt	cattctttta	ctaggcttag	960
cttccaaatt	atgctttata	gctgtataaa	catcgtgatt	atattcatct	acttagaaat	1020
tgttttatct	ttaaattaat	ttgcttagct	gtttgttttg	atgcttagat	tatgttctgt	1080
taatgggaat	ttaacatatt	taagaaacca	atatttaaaa	tgttgggtcta	ggtttttttc	1140
cttaacatat	attaccaggc	tttactgtat	ttcactcagc	cttaaatgtt	ataatatttt	1200
tggataacgg	ttattaattc	tttgagacct	tcgtatagcc	tataaaatgt	atgggagatg	1260
ttggtatttt	atgtgtataa	aagcaacaat	atcagcaact	tcgtgtttat	actgcacctt	1320
ggttgttgat	gtcaagtaaa	aaaaagattg	ttttgtaaca	cataaaaaaa	tgggaagaaac	1380
tgataccaca	cctaaggacc	aaagataaga	aagacttttt	gccaagaca	gtgaaagtaa	1440
ttataaaaaa	aagctttgac	cacttaccac	gtatctgaag	agatgagttc	atactatgat	1500
ttagaaagtg	gttcaattcc	cctgttggca	tatgattatt	tttactaaaa	ttaatacagc	1560
tctgtgggtc	ttccttagtg	ttttctttga	agccaatctg	tttttttttag	gacaccagcc	1620
tttggttttt	catctgttcg	agatgcctct	tctctgtctc	cttatcagat	agaaatggag	1680
tcattgtgctg	ctgcttcac	tagcagaggt	tgccctctgg	ctctgacact	ttttgtcagt	1740
tgtcttttagg	tggtcctgaa	tcttgggccc	ttttgattgt	gaatactgtg	tagcaggatc	1800
ttgagagtcc	ttgttcttac	ataggcattg	ctctagtttg	tctttggcaa	aaaaaaaaaa	1860
aaaaaaaaag	aaatatccag	ggaaccctgc	ccagactaat	actgttgggtg	gcataagaga	1920
atcaagccat	tctcaagaga	taacttcata	accagaattg	tctgttggct	agcagctgtc	1980
acagataggg	agggcacttg	ggatatgacc	tttctgtcca	ggtgattcac	agactagacc	2040
tttcttatcc	tcctcctaga	gttttgactt	gggactctag	tgtaaatgatg	atgagcccg	2100

gcatcaggtc	cttctgcact	ttgggtggaag	tctcccaggg	taggttttctt	atttgaaaca	2160
gtggaatcat	gtttccagtg	ataaagttta	atgacctcat	cctttttttt	ttttctcacc	2220
tgccatttgt	gtgtcttaga	tgggttttaa	ttgcatgaat	gtggctaata	tggttctcag	2280
aaattgggtca	gtatggcccr	acatagcttc	tgctctgtct	tactgactca	atacctttag	2340
gatttgtatc	agagtttggg	tactagtgtt	agtgggtggg	tcaccactac	ttaattggga	2400
gataatgaaa	ccaatcatgg	atgctgtttt	tattgggcat	gtcatctaag	agaggagaaa	2460
tagctgggtt	ttgggtctaa	ttatgaataa	ggactgattc	agaaaacgag	tttatggtag	2520
gtagactaaa	gtttcacatc	agactgtacc	attgtgattt	agacctatct	aaaattcaga	2580
gcataatcat	tggtgtacct	cagggtcacc	acccatgtat	tgggcttagt	caggattgac	2640
agatacattc	tcagctggcc	tgcatataa	aacatactgt	cattgagctt	aagctccgct	2700
tggtctgagg	tttcacctcc	atgtgtttca	ttgggtgcaa	agtggatctc	ttagtgtgtc	2760
acttaattct	ttctttttca	gaaagatagt	atgttccact	gtatatttgg	tcactcttag	2820
aaccttccct	cacattgttt	tttatgggac	ccatgaatgg	nttagccttt	cttttctatt	2880
gtagaaggaa	ataaatagga	gtaaaaagac	cattgtagta	aataagttca	aggggaactt	2940
gggaccagaa	accactgtta	tgtacaaaaa	aatggcaaat	tcaataaact	caaatttaaa	3000
ataattttta	aattaacagt	tatgataaat	tttatatttt	atacaaatag	attgcttaga	3060
atgggttctca	agaattataa	gagaaatgaa	ctcacagtac	aaaaatttta	taattactat	3120
acttgtgttt	tggttggggg	ctgggaaatg	tattttttaca	ttgtagccaa	tcattttata	3180
tttgtcaatt	taaatcttat	gggtcttttt	ttttttatct	ctcttgatgt	cagattttat	3240
agtcttttta	aataaatcca	tttaattaaa	acgttaaaaa	aaaaaaaaaa	aaaaaaaaaa	3300
ggcgggccgc	tcgcgaatct	agaactangn	cccacgcg			3338

&lt;210&gt; 37

&lt;211&gt; 1563

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 37

cggcacgagg	agaaaaggac	tcagaccttc	cagattgggt	ccagaaggca	gctgctgtca	60
gtgattttcc	ttacagacca	ttgcagccac	ctaccttgcg	ccatgggtccc	gagacagcat	120
ggcgtcttcc	cccacacgga	cctgcagggt	gcctcttagg	agctttctcg	ctgaaggagg	180
tcateccctct	tagtcacatc	tgagcagcac	tcacttaggc	ctcgcccccg	agggtcatgc	240
ggggggcacc	ggcgggggtc	ccggctctgt	ctgggtctggc	atctgtctgc	ctttgggcta	300
ccttttcggc	acagctccct	gggctctgtg	ctgctacaag	ctggactcca	gctccccctg	360
gctgcagtg	tgctcgcagt	gggtccagaaa	agaggctggg	cacagcggt	cctgggtcag	420
cagcatcctt	ggcgcagggt	ggcccaggag	ctccgtgcag	ggtgctgcca	gttgaccag	480
ccccggctgc	actcaatgtc	cgagaaccgg	gctgggtggg	gggctctttt	gacggagctt	540
tactccaagt	cttgctgaat	ttcctgagga	aaagcactga	tggtctcatg	gacacaaggg	600
aggcagaate	tcttgaagta	gaataaactg	cactcctttc	cagtcttctt	ttctcagctt	660
ctcctggacc	gccagctcct	ccatgcgcct	cagacactcc	caacgccgca	ctgcgggggt	720
tcctcccgcc	ctggcccatc	tcateccctt	tggtgttgga	tacagctccc	ctgtgtgcat	780
gtcgccctgt	ggcagatgct	cagggaactt	tcagactcca	ggatcactcc	gagtacactg	840
accacacagc	ctgctgcaca	aactgcagcg	ccagctaagg	accaggaatc	tgacattgtc	900
ggaggtgaag	ggatcctttg	tgacatagct	ttcctccagg	aggaccatcc	actgggtgtc	960
gggggagctt	ctgccccaa	tagcaggagg	gagctgtcaa	ggcgtggagt	tcacactcaa	1020
acactcccag	aggacggcac	tctgcatggg	acgccatcca	gctcctttga	ctgtggaatc	1080
aagtacatta	tcagctggcc	cctggctcct	ggctgtgacc	tccttctcgt	tgaactgagc	1140
cttgtgtgta	agggcgtatc	aagtgtgatg	ggcttcgcgc	ctgggtgacc	ccaaatgcaa	1200
ctccaggaag	catgggttac	caatgactaa	ggtaacgat	cagctgggcg	cggtgctgac	1260
gcctgtaatc	ccagcacttt	gggaggtcga	ggcgaatgga	tcacgaggtc	aggagtccaa	1320
gaccagcctg	gccaacatgg	tgaaaccccc	atctctacta	aaaatacaaa	aattagccgg	1380
gcgtggtggc	gggcgcctgt	agtcccagct	actcaagagg	ctgaggcagg	agaatcgctg	1440
gaactcggga	ggcgcagctt	gcagtgaagc	gagatctcgc	ccgctgcact	ccagcctggg	1500
cggcagagcg	agactccgct	tcaaaaaaaaa	aaaaaaaaaaa	aaaaaaaaaact	cgaggggggg	1560
ccc						1563

&lt;210&gt; 38

&lt;211&gt; 1048

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 38

cccggggtcga	cccacgcgtc	cgacaaaaaa	caaggtatgt	gtgtgccttg	ggatgctttt	60
ctgggctttac	cccatctgtg	tcttcataga	ttctctgtcc	tgccagccat	gtcttttggtc	120
cacaggagcc	acttcccact	tcaactctcc	cacaacctcc	ccactgttta	ctctcttcat	180
gccctgtgct	cttgccccaa	atcccttcac	ccaattggga	aaactggatg	atagataagt	240
agtttcttat	ttacatgcat	ttctctggaa	attaaagttg	ctttccagct	agtatttcct	300
cacgtcggaa	gcatgtgtaa	aataacacat	ggaaactatt	tccacctggc	acataagaaa	360
aaaagggctc	gcctatat	taactgactt	ctacactttg	catgttttag	tctgaccac	420
aatttttttt	tttggtatgt	tgaagtaaaa	aacgtgggta	caagatgact	aaagtttggc	480
ctacaatttc	ctgttttgac	aacaggaaga	attgtgcagt	gttttgcaac	aggactaatt	540
ctagattctc	actgccctta	aagataacta	gagggaaagg	gtcttctctt	tctttttcat	600
ttattaaaga	tacttacagg	ttcttcagaa	gaagtgtatg	gtccttttagg	taatgaaata	660
gggtgttgatg	gtgttatggg	tgatgatgta	actggtggcg	gctgtataaa	gtcttcatcc	720
ttaaacatca	gaactgagag	ttgggttact	gatttaaaaag	aaaatgtgta	ttaatctaca	780
gtccactacc	ttgactacca	ggaacttggtc	acaactacct	catatcctgg	cacaggattc	840
aaaaggcaaa	atctggaacc	taattacgtg	tagccaaaata	ttaagcttcc	tcttagatca	900
aacagtacaa	taaatattcc	ttattaagga	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	960
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1020
aaaaaaaaaa	aaaaaaaaag	gcggccgc				1048

&lt;210&gt; 39

&lt;211&gt; 1430

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 39

cggcacgagg	agcagctgag	tcccttccct	gtctttcact	cttctggcat	cggtgggtttt	60
acttcttcga	ttgaaccctg	cttctctgac	ccccctggga	ggccgccttc	ttcagggccc	120
tcccttctct	ccacgagctc	gctctgacag	ctgaggaact	ggcaagatcc	tgctaccag	180
aggggtgaatg	ggtatctttc	ccggaataat	cctaattttt	ctaaggggtga	agtttgcaac	240
ggcggccgtg	attgtaagcg	gacaccagaa	aagtaccact	gtaagtcatg	agatgtctgg	300
tctgaattgg	aaaccctttg	tatatggcgg	ccttgccctc	atcgtggctg	agtttgggac	360
tttccctgtg	gaccttacca	aaacacgact	tcaggttcaa	ggccaaagca	ttgatgccc	420
tttcaaagag	ataaaatata	gagggatgtt	ccatgcgctg	tttcgcatct	gtaaagagga	480
aggtgtattg	gctctctatt	caggaattgc	tccctgcgtt	ctaagacaag	catcatatgg	540
caccattaaa	attgggattt	accaaagctt	gaagcgctta	ttcgtagaac	gtttagaaga	600
tgaactctt	ttaattaata	tgatctgtgg	ggtagtgtca	ggagtgatat	cttccactat	660
agccaatccc	accgatgttc	ttaagattcg	aatgcaggct	caagggaagct	tgttccaagg	720
gagcatgatt	ggaagcttta	tcgatataata	ccaacaagaa	ggcaccagg	gtctgtggag	780
ggtaagtact	cttttctctg	tattatccta	cactctcagt	tcttacaatt	tgcagagaat	840
ttttttttat	ataaagacat	aaaatcgtga	attataatcc	aaaaactaag	gtaagaaact	900
cctcatctcc	cttgaaaggc	ccaaaactta	tcattggcct	tttatttctg	cataatgttt	960
ggggattata	taggtgggga	aagttattac	attattttgag	atggctgttt	cgatcatatt	1020
cacagtgaat	gtagttgttc	agtgtatttt	tttgcaagtt	ctgtactaac	acgatgatgt	1080
atgtctttgt	agtgtctatg	ttcaaaagct	gttaccggct	gtgtgctgtg	gctcatgcct	1140
gtagtccag	cactttggga	ggccaacgcg	ggtggatcac	ttgaggtcag	aagttcaaga	1200
ccagcctggc	caacctgggtg	aaaccccatc	tcaactagaa	atacaaaaat	tagccaggca	1260
tggtgggtgca	tggtctgtgt	cacagctact	caggaggctg	aggcgggaga	attgtttgag	1320
ccctggaggt	ggaggttgca	gtgagccaag	atcatgccac	tgcactccag	cctgggtgac	1380
ggggcgagac	tctgtctcaa	aaaaaaaaaa	aaaaaaaaaa	aaaactcgag		1430

&lt;210&gt; 40

&lt;211&gt; 2103

<212> DNA  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (2101)  
<223> n equals a,t,g, or c

<220>  
<221> SITE  
<222> (2102)  
<223> n equals a,t,g, or c

<400> 40  
ttcctcgtag cgagcctagt ggcggggtgtt tgcattgaaa cgtgagcgcg acccgacctt 60  
aaagagtggg gagcaaaggg aggacagagc cctttaaaac gaggcgggtg gtgcctgccc 120  
ctttaagggc ggggcgtccg gacgactgta tctgagcccc agactgcccc gagtttctgt 180  
cgcaggctgc gaggaagggc ccctaggctg ggtctgggtg cttggcggcg gcggcttctt 240  
ccccgctcgt cctccccggg cccagaggca cctcggtctt agtcatgctg agcagagtat 300  
ggaagcacct gactacgaat gctatccgtg cgagaacagc tattccacga gaggatccgc 360  
gagtgtatta tatcaacact tctgtttgca aactgtaca tcccttgcca catcttcttg 420  
acccgcttca agaagcctgc tgagttcacc acagggtgtc ctgggcccgg tctmtgagac 480  
agtgggtgatg ttgatgctcc tcaactctgt ggtgctaggt atgggtgtggg tggcatcagc 540  
cattgtggac aagaacaagg ccaacagaga gtcactctat gacttttggg agtactatct 600  
cccctacctc tactcatgca tctccttctt tggggttctg ctgctcctgg ctgctggaag 660  
acctggagga gcagctgtac tgctcagcct ttgaggaggc agccctgacc cgcaggatct 720  
gtaatcctac ttctgtctgg ctgcctttag acatggagct gctacacaga caggctcctg 780  
ctctgcagac acagagggtc ctgctgggta tgtggcttcg tagggcttgg gatacctggg 840  
tttccccaag gagagtagcc cctgggttcca ggtgcttgct gacagcctcc catccctgca 900  
cagagaagag gcggaaggct tcagcctgkc aacggaacct gggctacccc ctggctatgc 960  
tgtgcttgct ggtgctgacg ggcctgtctg tgctcattgt ggccatccac atcctggagc 1020  
tgctcatcga tgaggctgcc atgccccgag gcctgcaggg tacctcctta ggccaggtct 1080  
ccttctccaa gctgggctcc tttgggtgcc tcttatagct ctccactctt ccggagcctg cggcccagat 1200  
tggtgtcctc agttgtgggc ttctatatgt ctccactctt ccggagcctg cggcccagat 1200  
ggcacgacac tgccatgacg cagataattg ggaactgtgt ctgtctcctg gtcctaagct 1260  
cagcacttcc tgtcttctct cgaacctctg ggctcactcg ctttgacctg ctgggtgact 1320  
ttggagcgtt caactggctg ggcaatttct acattgtgtt cctctacaac gcagcctttg 1380  
caggcctcac cacactctgt ctggtgaaga ccttactgct agctgtgcgg gcagagctga 1440  
tccgggcctt tgggctggac agactgcccg tgcccgtctc cggtttcccc caggcatcta 1500  
ggaagacca gcaccagtga cctccagctg ggggtgggaa ggaaaaaact ggacactgcc 1560  
atctgctgcc taggcctgga gggaagccca aggtactctg gacctcagga cctggaatct 1620  
gagaggggtg gtggcagagg ggagcagagc catctgact attgcataat ctgagccaga 1680  
gtttgggacc aggacctcct gcttttccat acttaactgt ggcctcagca tggggtaggg 1740  
ctgggtgact ggtctagacc cctgatccca aatctgttta cacatcaatc tgcctcactg 1800  
ctgttctggg ccatccccat agccatgttt acatgatttg atgtgcaata ggggtgggta 1860  
ggggcagggg aaggactggg ccaggggcagg ctggggagat agattgtctc ccttgccctt 1920  
ggcccagcag agcctaagca ctgtgctatc ctggaggggc tttggaccac ctgaaagacc 1980  
aaggggatag ggaggaggag gcttcagcca tcagcaataa agttgatccc agggtttctt 2040  
ttgttttttt aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2100  
nna 2103

<210> 41  
<211> 2349  
<212> DNA  
<213> Homo sapiens

<400> 41  
tcgacccacg cgtccgtcta tttctgtgtc tgtatgtatg ggatgtgtct gattgtgtgt 60

ctgtggtgta	tttctgtgtc	tgtatgtatg	gggtgtgtgt	gattgtgtgt	gtgtctgtgt	120
tatctctgta	tgtatggggt	gtgtctgtgt	gtgattgtgt	gtgtgtctgt	tgtatctctg	180
tgtctgtatg	tatggggtgt	gtctgtgtgt	gattgtgtgt	ctgtgtttat	gtgtgtgtgt	240
ctttgtgtga	ttttttgtgt	atatggcaaa	cccagaacag	aacattatca	ctcacctcac	300
ctggcaaaac	aaaaagcttt	cagagagatg	tgtggccgtc	atgatgtctc	agctgcaggc	360
atttttcaaa	gttatgtttg	atgcagctgg	cacctttctt	ctcggggaca	aatgtgatct	420
ttgatcattg	agttgagaat	gagtttactc	aataatttta	aaatgcctcg	gtgtgagagg	480
atgggctgct	agtaaacatt	tgtggcaaat	atgcgttgcc	ttttctctct	tcaagtgcac	540
ggggcaggct	gcttctctgc	acgtccatct	gtgggtcagt	ctgtctctgc	acctcccttg	600
ccccagggct	tagtgccctg	tgtggatcac	ggctgcccc	gggatgctgg	ggaaggattc	660
agcccaacaa	atagattcgt	tgacgtcctc	tcaacgaatg	tgtgctttta	aaaatgtctt	720
taattttgca	tcccatttca	tacaacaggt	gcacatttaa	aacattttat	tttaaacagc	780
tctgtcttct	cagcactgca	gactatttct	cgctcacttg	acaggcagac	gttgggacta	840
tgttcttttc	ctgccaagaa	tccatcagcc	ccccaatgcc	tgcacctcga	gtccagactc	900
cttagcggtg	cgttgggacg	taattccccg	ttagcatccc	cagccccctc	tccccctgtg	960
agccagcctg	ggttcttttc	tttttctctt	cttccctgtt	ttcttctctt	cttttttggg	1020
tttgtaataa	gtaatatata	gtcttgtggg	taagaaatca	aaataataaa	gaagatacgg	1080
cttgaggatt	cttgctctgc	actggctcct	ccgccaggac	ctggtgcgtc	ccccaaagagg	1140
gccttgccat	tgcgagctg	agacacatgt	tcacactttc	ctcctttttt	gtacaaaagg	1200
tagcacatgg	taaatagcgt	ctgcacgctt	ctccccctgc	tctgtcttcg	ggactactcc	1260
acatccatga	agaagcagca	tccttattct	cttgccacagt	gggcagtgcc	tacactgtat	1320
tcactcgtct	ctgctgatgg	gcactattgt	agtggtgcaa	ttagtaacct	agtatacatg	1380
cccttcagcg	tgtgtactgg	tgtatggatc	tcacattccc	agaggctgag	tgaagaataa	1440
aaggcttttt	gacagtattc	ctcaggtact	cacacactcg	ctctgtgagc	cacgtgatca	1500
tgggagcggc	tccctgcaga	ggtgtctgat	gggagaggct	gatgtccttg	tctgatgggt	1560
gctccaacct	gagaattggt	agctgctctc	ccaatgcaga	ctttcatcac	ttgcatctca	1620
tgcgtgagat	tgaactgctt	ctcatttatg	tttttatctt	tgtttgcca	tatttttagct	1680
gggttgctaa	tcttttaaat	ttttctaggc	gcttttcata	aattagggag	tttagctgta	1740
gtgatgtgag	ttgaaagtat	tttttcaaat	cgtcattgtc	ttgactttcc	tttttttttt	1800
tcttttgcca	cttaaaatat	gtagttgaat	ctcagtatct	gggggcttct	ggatttttgag	1860
tcgtagttag	aaaggtcttc	ctattctgag	gttatcacagg	aattcaccca	tttttttctc	1920
aatattttat	ttccatttaa	acttttgacc	cacttggatt	gattttgttg	tatgggtgtca	1980
agtatgggtc	caagtattgt	tttctgacct	ggggatccca	acatcgttca	tcaaaatctc	2040
catgttggcc	gggcgtgggt	gctcaagcct	gtaatcccag	cactttggga	ggctgaggtc	2100
ggtggatcac	aaggtcagga	gatcgagacc	accctggcta	acacggtgaa	atcctatctc	2160
tactaaaaat	acaacaaatt	agccgggtgt	ggtggcaggc	acctgtagtc	ccagctactc	2220
aggaggctga	ggcaggagaa	tcgcttgaac	ccgggaagcg	gaggttgag	tgagctgaga	2280
ttgcgccact	gcactccagc	ttgggtgaca	gagcaagact	ctgtctcaca	aaaaaaaaaa	2340
aaaaagggc						2349

&lt;210&gt; 42

&lt;211&gt; 1559

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 42

attcatgcca	aaacataggc	tttcagtgcc	tattacatat	ggctttcagc	tctctctact	60
gagggatgta	ggagtttatt	tctgaggtct	gagcctcttt	tcctttactt	cctttactct	120
ttcctaagcc	ttctttataa	aaactatgca	tgttctattg	ttttcttttt	tgattccctt	180
tctttttatta	ttcccagtag	gagtgcattg	taattctcat	atgttagaaa	ggcaggtctc	240
ctggttgaag	aaaagatcca	ccaagcaag	tcagcagttt	aataaatttt	tgagggggat	300
ctcaaattgt	ggaaggattg	ttatataaga	caaccaaatt	atgacatgag	acaataaatt	360
ctataggaat	tatggaggaa	taattagcta	tttattttct	tggttaggga	agagatatta	420
ttagtgttag	aagtaattac	taacttctac	attttttatt	gtggaaatca	aaaatatata	480
tatgaaaata	aaatgttata	attgacttca	gtgtcccata	aaccagcttc	caacaattac	540
caaattgtga	ccaatccttt	acacacatgc	acagggtgcc	ctccagtatc	tgtggggcat	600
tggttctagg	accacttatg	gataccaaca	tctatggatg	ctcaagtccc	tgatataaaa	660
tgggtggacta	tatgcatata	acccgtgtac	gtccccgtatt	aaccaaattcc	ttcctagatc	720



acttacaaca	cgtaacacaa	tctaaatgcc	acgttaacaa	ccgtcatact	gtattaagcg	780
aataaccccg	ccaaaaatgt	acttctctca	gtccagacgc	cttttttttt	tgtgtgtgga	840
atatcccagc	ccccaaagtc	ccttcaaccc	cctggcacat	aggaggctga	ctgcgtgtgt	900
gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgcatac	agacacacat	atctctgaaa	960
tgtaaatatt	ctctctttta	aaaaattatt	atcacagcta	aacaaattac	cagtaattct	1020
cttatccctca	tatacccggt	gttcagattt	tctagattgg	ctcctaattt	ttttacagat	1080
tatttgaatc	tgattcaatt	catgtactgt	aatgtttgat	aacttaagta	ccctttatag	1140
gttcccttta	ccacttcttt	attaaattcc	ttgtaatttg	ttgtactaaa	tggattgtct	1200
tcctagaatt	tcctgtagtc	tgaattatgt	ggtaattgtt	cacatgttcc	agtgtcctct	1260
tatttcctgt	gagttggtag	ttagatctag	aagcttgatt	aaattcagat	tttctctctt	1320
tagatcatca	acttttagatc	atcaacttgg	atcatttggt	tcattttgct	tttgatatgt	1380
tgttttttag	aattacctct	taaaattttg	atttaatttt	ataatcatgt	aaaatgttta	1440
taaatttcca	aattcagatc	agcaaaacac	aataaaatct	attcagagaa	ggcaaaaaaa	1500
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1559

&lt;210&gt; 43

&lt;211&gt; 1766

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 43

cggcacgagg	agcactgaag	tattcactac	atgaagtata	ttttgcaactg	tggacacaaa	60
ttagaaaaat	tgcaagtagt	ggtatattgt	aattggcatg	cactatatga	gcagagtcaa	120
tgtgtctcct	tgtagaatat	tctctgatga	tactcactat	tatccccctc	ctgctaagct	180
ttgttctgtg	tctgaagggc	ataaagcatg	gaaactacat	ttttcagact	ccattaccag	240
aaggatatgg	ttggatttca	gcaatgagtg	ggctttgcat	aaaatttggg	agacgaaaga	300
gaagaaaaac	ctggctgctg	caggttggaa	cactggcaac	aatagatacg	gagtttgcaa	360
gaagctgcta	agcttctctc	ggaaaattat	ttgtttcaat	atttctggca	tatgggatca	420
tcaattattg	ttttcagtg	ttctggctga	aaattgggtt	aattcttcta	tctgagaatt	480
gttcatttct	gtgcttcagg	aaactaagac	catcactggc	agtttttggt	gagggatcct	540
tgtgcattca	tttttcttta	aaacagcctt	cctaactttt	actccccag	cctctatggg	600
tgtgtaagtc	tttaattctt	agagttacat	ttctcttact	cagtatatcc	tagtgccggc	660
tctgttttcc	agaccacaac	ctgactgata	tagtctccat	gtgtttcaga	tgggtgggata	720
gtttggatat	ttgtccctgc	ccaaatttca	tgttgaactc	taatccccag	tgctgtaggt	780
ggggcccaat	gggaggtgtt	cggatcatgg	gggcagatcc	ctcacgggtt	gggtgtgtct	840
tcgtgatagt	gagttcttgt	aagatctggg	catttttaaag	tgtttggcac	ttgtaccatt	900
tcactgtgtc	ttgctcctgc	tttcaccatg	tgaagtgcct	gctccagctt	cacttttccac	960
catgattgta	aacttcctga	ggcctcccta	gaagccaagc	agatgccatc	accaggggtc	1020
ctgtaaagca	tgcagaactg	ttagccaatt	atacctcttt	tctttacaaa	ttaaaaacct	1080
cttttcttta	caaaatggaa	agaataaagg	tatttcttta	tagcaatgca	agaacggcct	1140
aatacagatg	gctctgccat	tagtgagaaa	attgagacgc	tttctctaga	tggcaaaaaa	1200
gttgtaaaaa	taaaaggaaa	ttattaatat	accgtctatt	tgatcatttt	actaagttaa	1260
gcatattatt	aagaagacaa	gcataagttt	aacacaattt	ggcaatgaat	aaaattgaa	1320
gagagagagc	atatgttggc	tttgctctgt	gaaactcaaa	tgaaatgggc	acctgttcta	1380
gcagctcatg	aaaaattctg	cattgtttat	tatgtgtcag	gatcaacctt	aaattcagtt	1440
ataaaaagtt	tgatgattac	aaaaaaaaagg	gaagcactaa	gtaatatagg	tacagagagg	1500
gaagagtgtc	aaatagaatt	ttcaatctgt	gtataaggat	acttaagcat	tttttaagga	1560
aagcagaaa	aagcatgaga	aagctctaca	tgacatctat	gtcaatataa	caagctggat	1620
atthagagaa	gaaactcttg	attaaatact	ttttatgata	tgaacacaca	catataatat	1680
gacatgactg	tgttcatgga	acataaagaa	atttctctga	ccaaagagaa	ctggaaaaaa	1740
aaaaaaaaaa	actcgagggg	gggccc				1766

&lt;210&gt; 44

&lt;211&gt; 2572

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

<220>  
<221> SITE  
<222> (2527)  
<223> n equals a,t,g, or c

<400> 44

aattcggcac	gagtctggac	cttcttagmt	tgcttgatat	caggttggtt	ttgtagccat	60
tttgcttcac	agtcacctgg	aatgccggga	gcccctgctc	atcccgatcc	tctccttgta	120
catgggcgca	cttgctgcgt	gcaccacctt	gtgcctgggc	tactacaaga	acattcacga	180
catcatccct	gacagaagtg	gcccggagct	gggggggagat	gcaacaataa	gaaagatgct	240
gagcttctgg	tggccttttg	ctctaattct	ggccacacag	agaatcagtc	ggcctattgt	300
caacctcttt	gtttcccggt	accttggttg	cagttctgca	gccacagagg	cagtggcgat	360
tttgacagcc	acataccctg	tggtcacatg	ccatacggct	ggttgacgga	aatccgtgct	420
gtgtatcctg	ctttcgacaa	gaataacccc	agcaacaaac	tggtgagcac	gagcaacaca	480
gtcacggcag	cccacatcaa	gaagttcacc	tctgtctgca	tggtctgtgc	actcacgctc	540
tggttctgta	tggtttggac	acccaacgtg	tctgagaaaa	tcttgataga	catcatcgga	600
gtggactttg	cctttgcaga	actctgtggt	gttcccttgc	ggatcttctc	cttcttccca	660
gttccagtc	cagtgggggc	gcattctcacc	gggtgggtga	tgacactgaa	gaaaaccttc	720
gtccttgccc	ccagctctgt	gctgcggtat	atcgctctca	tgcgagcctt	cgtgggtccta	780
ccctacctgg	gggtgcacgg	tgcgacctgt	ggcggtgggt	ccctcctggc	gggctttgtg	840
ggagaatcca	ccatggctgc	catcgctgct	tgctatgtct	accggaagca	gaaaaagaag	900
atggagaatg	agtcggccac	ggagggggaa	gactctgcca	tgacagacat	gcctccgaca	960
gaggaggtga	cagacatcgt	ggaaatgaga	gaggagaatg	aataaggcac	gggacgcat	1020
gggcactgca	gggacagtca	gtcaggatga	cactctggca	tcatctcttc	cctctcccat	1080
cgtattttgt	tccctttttt	ttgttttgtt	ttggtaatga	aagaggcctt	gattttaaagg	1140
tttcgtgtca	attctcttagc	atactgggta	tgctcacact	gacgggggga	cctagtgaat	1200
ggtctttact	gttgctatgt	aaaaacaaac	gaaacaactg	acttcatacc	cctgcctcac	1260
gaaaacccaa	aagacacagc	tgccctcacgg	ttgacgttgt	gtcctcctcc	cctggacaat	1320
ctcctcttgg	aaccaaagga	ctgcagctgt	gccatcgctc	ctcggtcacc	ctgcacagca	1380
ggccacagac	tctcctgtcc	cccttcacgt	ctcttaagaa	tcaacagggt	aaaactcggc	1440
ttcctttgat	ttgcttccca	gtcacatggc	cgtacaaaga	gatggagccc	cgggtggcctc	1500
ttaaatttcc	cttcggccac	ggagttcgaa	accatctact	ccacacatgc	aggaggcggtg	1560
tggcacgctg	cagcccggag	tccccgttca	cactgaggaa	cggagacctg	tgaccacagc	1620
aggctgacag	atggacagaa	tctcccgtag	aaagggtttg	tttgaaatgc	cccggggggca	1680
gcaaactgac	atggttgaat	gatagcattt	cactctgcgt	tctcctagat	ctgagcaagc	1740
tgctcagttct	cacccccacc	gtgtatatac	atgagctaac	ttttttaaat	tgtcacaaaa	1800
gcgcattctcc	agattccaga	ccctgccgca	tgacttttcc	tgaaggcttg	cttttccctc	1860
gccttttctg	aaggtcgcgt	tagagcgagt	cacatggagc	atcctaactt	tgcatttttag	1920
tttttacagt	gaactgaagc	tttaagtctc	atccagcatt	ctaattgccag	gttgctgtag	1980
ggtaactttt	gaagtagata	tattacctgg	ttctgctatc	ccttagtcata	actctgcggt	2040
acaggtaatt	gagaatgtac	tacggtaact	ccctcccaca	ccatacgata	aagcaagaca	2100
ttttataacg	ataccagagt	cactatgtgg	tctcctctga	aataacgcat	tcgaaatcca	2160
tgacgtgcag	tatatTTTTT	taagtTTTgg	aaagcagggt	ttttccttta	aaaaaattat	2220
agacacgggt	cactaaattg	atttagtcag	aattcctaga	ctgaaagaac	ctaaacaaaa	2280
aaatatTTTa	aagatataaa	tatatgctgt	atatgttatg	taattttatt	taggctataa	2340
tacatttctt	atttttgcgt	tttcaataaa	atgtctctaa	tacaatacgg	tgattgcttg	2400
tgtgctcaac	atacctgcag	ttgaaacgta	ttgtatcaat	gaacattgta	ccttattggc	2460
agcagtttta	taaagtcggt	catttgcat	tgaatgtaag	gctcagtaaa	tgacagaact	2520
atttttncat	tatgggtaac	tggggggaata	aatgggggtca	ctgggagtag	gg	2572

<210> 45  
<211> 526  
<212> DNA  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (66)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (106)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (484)

<223> n equals a,t,g, or c

<400> 45

ctctgacagc	ttcctctttg	gccaaagccct	gcctctgtac	agcctcgagt	ggacagccag	60
aggctnagac	tggagcccag	agcccaagat	ggagccccag	ctgggnccctg	aggctgccgc	120
cctccgccct	ggctggctgg	ccctgctgct	gtgggtctca	gccctgagct	gttctttctc	180
cttgccagct	tcttcccttt	cttctctggt	gccccaaagtc	agaaccagct	acaatttttg	240
aaggactttc	ctcgggtcttg	ataaatgcaa	tgcctgcac	gggacatcta	tttgcaagaa	300
gttcttttaa	gaagaaataa	gatctgacaa	ctggctggct	tcccaccttg	ggactgcctc	360
ccgattccct	ttgstttctt	atccttgcaa	attactccar	atgattycca	aaatctggsg	420
sccttgtgga	ratcttttaa	ctggtcagca	awtwcacaac	gaaatctcca	aacaggaaat	480
cttntgcctc	ctgcatccac	ccccaaagaa	cttgcacatt	gacgtt		526

<210> 46

<211> 1032

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (974)

<223> n equals a,t,g, or c

<400> 46

gtaaaattgt	caaacattta	tttttaaaaa	taaaatgtgg	tgggagggtta	tgggtggagt	60
tattgggtcc	tggctgtcgc	cactgagcat	tgtgaatgt	tgtcatgact	tgtggacttc	120
tcagagctgc	gagcatgcgg	gtgctctttg	tggagatttg	ctctgtgcct	gtaggaagg	180
gggtgtgtgg	tgtgcattgc	agcaacattg	gtggaacaga	tgtgtgtgcc	cccatgctgt	240
catcagggtt	cattgcacag	gagcctcata	cacactacag	aaaatatgct	cttgtaaccc	300
caagtccatg	gggaggcatc	cacacagatg	gcagcagatc	aggaaatgct	ctcagccagt	360
cctcaggggc	agtagggcag	ccttcacatg	ggtgcgcttg	gcagccctga	actttatctc	420
ttcttttctg	tgtatttctt	taatatctta	ttcagcattt	tagtttcttt	agttggatga	480
gttggttcagt	gtctttgtac	tgtacagtgc	tcctgtgctg	ttggaatgga	cgtcattggc	540
tatgttactc	ttttgaaata	caaatttctc	tttagaagtt	ttggaaaatc	tgtttagagga	600
ggagttctga	gaatatgaga	agacttcttg	ccacatagtt	gtgcctgccc	gatgtatttt	660
tccaaaaaaa	ttgaagggtga	acctgttgga	agctcagttc	tgagtttaat	tgactacttt	720
gcagagacac	aacactctct	ttsttagcatg	tcccttggaa	gaaaataggc	tgggtgcttg	780
gtggtgatca	gaactggcag	ttattctagg	cctaaaacag	tttggcaata	aaataagaaa	840
aagggggcac	atgtgtgtgc	tcattgcttta	atccccacac	tttgagaggc	tgaggcaggg	900
ggattgcgtg	aagccagcct	gggcaacata	aaatttataa	attagccaga	catggtggca	960
tgtgcctgtg	ggtncagcaa	cctgggtggc	taagaccgaa	ggatcacttg	agcccaggaa	1020
gtcaaggctg	ta					1032

<210> 47

<211> 2680

<212> DNA

<213> Homo sapiens

&lt;400&gt; 47

ggcagcagtc	ttacagcaag	tacttaacaa	acttagattt	ccataatcag	atztatctca	60
aaccgcaaaa	aaaagctgat	tttattatgt	taaactgata	ggagatTTTT	atgagagatg	120
aaagtaagcg	atTTtaactt	TTtaattttt	ttaatttttg	ctctgtttct	cacacttgag	180
gcttttctga	aattcaccaa	aagagttttg	gcagttgtgg	ggaatttacc	agaaccccc	240
attattaaaa	caattgggtt	cttatattaa	aatcatcaga	caggttttgt	gtacacagac	300
atztatgtta	agccactgga	tagaactaaa	ctgtgtgtag	gaacctgttt	tctctaagaa	360
aggactcctt	tgagtTctgt	aggatatgtg	cttaaagggt	ttctgtgtca	agctatttta	420
tgaaataaaa	ttttacaaac	aatgaagctt	cattttcaca	cgttgtatTT	gtttaatgca	480
gacttgccct	atacgagttt	ttgaaattaa	gggttttttc	cgtaaaacta	gcttaggagt	540
ttaatgagtc	actgtgtctc	tcacctctgt	aggatgagca	cagagatcac	atTTgagaag	600
cagggtgtgg	tgTTTTctgc	tggtcacgtc	tcctctgcct	ctcctctggc	tcagcagtta	660
cttttctca	gtcatggggg	ctgcattcag	ctgcgagcc	agcattaggg	ccaagctctc	720
ccgcttctgt	gacgttttat	tacacgtact	ctttatcagt	agtagctgcc	attccctgag	780
tgttccttcc	accgtgcact	gagaagggtc	gaggatttcc	gaacatcaac	ctcctctaac	840
cctctgtatg	ggaattagat	ttttatcatc	ctgactgttg	gggctttcta	gagctcaggc	900
aacttgcccc	tgggcgggat	ggagcctgcc	tggggggagt	gtgaatgagg	cttctgttgt	960
ctgtggcttg	gctctctgtg	gttgaaact	tcagaagcag	tagattccag	gtgcaacatt	1020
tggtgcccgg	acaagtTgta	aaacctgacc	ttggggaggg	ccttcaactgg	cagccagtgt	1080
ccaccctga	gcccggTgag	cacagacgtt	agagccaggg	ccggcgggct	cagaggttgg	1140
agccggggac	cagtgggctc	agaggTcgga	gccggggacc	agtgggctca	gaggTcgag	1200
ccatgtccct	gccgtgaagc	tctgagggtt	ggtggTgaca	tcggggcgat	tgtggcatcg	1260
cgattgtggT	gtcttggtcg	cctttggccg	gcgctgccac	tgccttggtt	tcatgacgta	1320
ggcaccacat	ggtcctgtgt	tgtctcactt	acggctccac	ttctggcttt	tttctacttg	1380
ctcacccttg	agataaaactt	ctattttcatt	ttcatttcgg	tttatatgtg	ggttttcttc	1440
caggTctgat	gttaagccta	taatatTgca	atgtgatgtt	ttgaagttaa	ggtgtaatatg	1500
agccagtga	ccaagggTtc	acaccccagt	gaaatacaaaa	tattcagaat	tgagccactg	1560
tgttgccata	ctgattatgt	aatgtgtgat	taacaagtat	aatgtgtcac	tttcaacatc	1620
agtttcatgc	caaagtTgca	ttttattaga	ttatttgTga	gttcaactttg	ggcccaaagg	1680
ctcgtgtcta	cataataata	acttatgatt	tttctttttg	tctttgtttt	atTTTTgttt	1740
ttgtgttttt	tgctttctag	accatgccag	agtaatctca	gctttcttta	gttactggat	1800
cacacatac	cttctctgaga	agagcagtga	ctaaaaTgga	atatctcttt	aagaacagct	1860
cctctttaac	aaaaaaactt	aaaagacaaa	tgtgagatgg	gcttagagtt	agttctctgg	1920
gaacttgaaa	gacatttatg	ccatattatt	tattcacgtg	tttgttcctg	gtgggcaaga	1980
tgccatctga	ggcttcagat	gagaaattgg	ggtaaaaTgg	aaatttttca	cttattTgca	2040
attatatata	tcttgaatta	ctacataaaa	cttgattctg	tttctctact	tattgtaaaa	2100
attgaaaatg	gacattctgt	taagtTaaat	gtatagtttg	aagctcatat	atTTTTatga	2160
agttttgaat	cacctgttat	ctgaaagtct	ctgctttTaa	aatgctttct	gggtattaaa	2220
atgttctagt	ttaagtagtT	tgaatatagt	tgagtTTTTt	ttctcttctc	tactttgtga	2280
atcatatcag	gtacctgttt	ttcctgtttc	gattttcttt	tctgtgatag	aagcagtcgt	2340
cagttcttgg	tattactaag	tgTtaaaagc	atcagtcagg	ccgggtgcgg	tggctcacgc	2400
ctgtaatcct	ggcactttgg	gagaccgagg	caggcggatc	acaacgtcag	gagatcgaga	2460
gcatcctggc	taacacggtg	aaaccccatc	tctattaaaa	atacaaaaaa	ttagccgggc	2520
gtggTggcgg	gcgcctgtag	tcccagctac	tctggagggt	gaggcaagag	aatggcatga	2580
acctgggagg	cagagctTgc	agtgagctga	gatcgcacca	ctgcactgca	gcctgggcga	2640
cagagcgaga	ctccagctca	aaataaaaaa	aaaaaaaaaa			2680

&lt;210&gt; 48

&lt;211&gt; 1730

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 48

cccacgcgtc	cgggggcctg	tggcggagtc	caagaggaac	taggcgcctt	tcgggatggT	60
gtggagcgct	gctcctgccc	cctgttgccT	cctcggggTc	ctggggctgg	tccaggTgtt	120
gggggcccac	gccgtgggTc	cctggacggc	ttcagcgtgt	ttgggggcag	ctcaggTctc	180
gccctgcagg	ccctgcaagg	agagctctct	gaggTtatTc	tCagcttcag	ctccctcaat	240

gactcactga atgagctcca gaccactgtg gagggccagg gcgctgatct ggctgacctg	300
ggggcaacca aggaccgtat cattttctgag attaacaggc tgcagcagga ggccacagag	360
catgctacag agagtgaaga gcgcttccga ggcctagagg agggacaagc acaggccggc	420
cagtgcacca gcttagaggg gcgattgggc cgtcttgagg gtgtctgtga acggttgga	480
actgtggctg ggggactgca gggcctgcgc gagggccttt ccagacacgt ggctgggctc	540
tgggctgggc tccgggaaac caacaccacc agccagatgc aggcagccct gctggagaag	600
ctggctgggg gacaggcggg cctgggcagg cggctgggtg cccttaacag ctccctgcag	660
ctcctggagg accgtctgca ccagctcagc ctgaaggacc tcaactgggc tgcaggagag	720
gctgggcccc cagggcctcc tgggctgcag ggacccccag gccctgctgg acctccagga	780
tcaccaggca aggacgggca agaggggcccc atcggggccac caggtcctca aggtgaacag	840
ggagtggagg gggcaccagc agccctgtg ccccaagtgg cattttcagc tgctctgagt	900
ttgccccggg ctgaaccagg cacggttccc ttgcacagag tcctgctcaa tgatggaggc	960
tattatgata cagagacagg cgtgttcaca gtccgactg gctggacgct acttgctgag	1020
cgcggtgctg actgggcacc ggcacgagaa agtggaggcc gtgctgtggc cgctccaacc	1080
agggcggtgg cggcgtagac tccgggtggc tacgagcctg agggcctgga gaataagccg	1140
gtggccgaga gccagcccag cccgggcacc ctgggcgtct tcagcctcat cctgccgctg	1200
caggccgggg acacgggtctg cgtcgacctg gtcattggggc agctggcgca ctcgaggag	1260
ccgctcacca tcttcagcgg ggccctgctc tatggggacc cagagcttga acacgcgtag	1320
actgggggtc cgcccgacgt gtctacgtcg gctgaagaga cagcgggggc ggcgggctcc	1380
tggggtctcg cctgagacgg ggcacctagc cctgggcgag cgccgcaccc gggcccgag	1440
cggcaccgcg cccagagcgg cctctcccca cgcccggggc gcgccggctc agggagcctc	1500
ggggccgccc atgcagactt ttggcctggc gcgatcccc aagaaccct ccaggccgg	1560
cctgcggagg agccgatcct cgcaccctcc gtcctctcca ctggccctcc aggtcgattc	1620
cctgggctcc aggtcctccc gcgcggggcgc cgcccaccgc cataactaac gatcgaggaa	1680
taaagacact tggtttttct aaaaaaaaaa aaaaaaaaaa aaggggccgg	1730

&lt;210&gt; 49

&lt;211&gt; 1275

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 49

ggcagagacc agctcctctt ccaagatgtc ctccttaaac tttgccagtg aaatggagaa	60
gtactgtttg ggggaataata tgctttcaag gttttgctta tttttgataa tgctcctgca	120
catccttctt tttttggtta tcttcattca acgtcacact gttgtctctc tctccaagca	180
ccaceeettt gttccaacta atggatcaaa gagttatagc agcttttaag gcctgattac	240
tacctgagga ggacctttgc ctaggctatt gctgcaata agaaagatcc tttttaaaca	300
cactgggtga attctggaag gattacaaca tctatggcag catcaacaac cttgggttaag	360
ttgggggtga agtcaccaag gagtgtgtaa atgctatctg ggaagagata ttgaagacat	420
ttgtccatga cttcaaaaca ttgctaagga tgaggatgtt gcaaaaaaaaa aatattcaca	480
aggcttttagt tgaaatggca aacaacttta agctggatgt ggaggaggat gatattgagg	540
aggtcacaga tgtggctact gaggaattga ctaatgacga gttactggaa ccagaacaga	600
aatgcatagc aagaggaaag gaaactacag gagaaaaaaaa gaaaagcccc caaggatatt	660
cacagtgatg ggtttaggca gaagcttttg cagacctcaa caaactcctt aaaacgtttg	720
aaaatgttga ctccaacatg aaatgatttt cattagtaga gaggaatgtt catcgtgcat	780
tatctattta caagcaaadc tatgatgaat aaaagaaaca aaccaaaca accaccatga	840
acacattttc gaaaagagtg acatctcaag aagagcctca ggcaggctct tcaggaggta	900
tcccagaaga agaaggcatt gttatcacag gacatgacag ctccctgggt gttattgccc	960
ttaaaggcct tccagtagga taagggtgtg agatggaagg tgccgatttt tatgatcctg	1020
accctacata ggtctaggct aatgggtaat ggtgtgtttg tgtcttagtt tttaaagaaa	1080
aatgtaaaaa gtaaaaaata aaaataacat taatataggt aaaagtttat agaataagga	1140
tataaggaga gaaaaaatt atcttctctc gtattttgtt ttcaagctgt gttactacaa	1200
aagaataaaa aagttaagag aaattgtttt taaagtaaaa tagttacaat aatctaaaaa	1260
aaaaaaaaaa aaaaaa	1275

&lt;210&gt; 50

&lt;211&gt; 1762

<212> DNA  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (447)  
<223> n equals a,t,g, or c

<400> 50

ggtggcctag	agatgctgct	gcgcggttg	cagttgtcgc	gcacgcctct	gcccgcacgc	60
ccgctccacc	gccgtagcgc	ccgagtgtcg	gggggcgcac	ccgagtcggg	ccatgaggcc	120
gggaaccgcg	ctacaggccg	tgctgctggc	cgtgctgctg	gtggggctgc	gggcccgcac	180
gggtcgccctg	ctgagtgggc	agccagtctg	ccggggaggg	acacagaggc	cttggtataa	240
agtcattttac	ttccatgata	cttctcgaag	actgaacttt	gaggaagcca	aagaagcctg	300
caggaggggat	ggaggccagc	tagtcagcat	cgagtctgaa	gatgaacaga	aactgataga	360
aaaktttcatt	gaaaacctct	tgccatctga	tggtgacttc	tggattgggc	tcaggaggcg	420
tgaggagaaa	caaagcaata	gcacagnctg	ccaggacctt	tatgcttgga	ctgatggcag	480
catatcacaa	tttaggaact	ggtatgtgga	tgagccgtcc	tgccggcagcg	aggtctgcgt	540
ggtcatgtac	catcagccat	cggcacccgc	tggcatcgga	ggccccctaca	tgttccagt	600
gaatgatgac	cggtgcaaca	tgaagaacaa	tttcattttgc	aaatattctg	atgagaaaacc	660
agcagttcct	tctagagaag	ctgaagggtga	ggaaacagag	ctgacaacac	ctgtacttcc	720
agaagaaaca	caggaagaag	atgccaaaaa	aacattttaa	gaaagtagag	aagctgcctt	780
gaatctggcc	tacatcctaa	tccccagcat	tcccccttct	ctcctccttg	tggtcaccac	840
agttgtatgt	tgggttttga	tctgtagaaa	aagaaaacgg	gagcagccag	accctagcac	900
aaagaagcaa	cacaccatct	ggccctctcc	tcaccaggga	aacagcccgg	acctagaggt	960
ctacaatgtc	ataagaaaac	aaagcgaagc	tgacttagct	gagacccggc	cagacctgaa	1020
gaatattttca	ttccgagtgt	gttcgggaga	agccactccc	gatgacatgt	cttgtgacta	1080
tgacaacatg	gctgtgaacc	catcagaaaag	tgggtttgtg	actctgggtga	gcgtggagag	1140
tggattttgtg	accaatgaca	tttatgagtt	ctccccagac	caaattgggga	ggagtaagga	1200
gtctggatgg	gtggaaaatg	aaatatatgg	ttattaggac	atataaaaaa	ctgaaactga	1260
caacaatgga	aaagaaatga	taagcaaaat	cctcttattt	tctataagga	aaatacacag	1320
aaggctctatg	aacaagctta	gatcagggtcc	tgtggatgag	catgtggtcc	ccacgacctc	1380
ctgttggagg	cccacgtttt	ggctgtatcc	tttatcccag	ccagtcaccc	agctcgacct	1440
tatgagaagg	taccttgccc	aggtctggca	catagtagag	tctcaataaa	tgtcacttgg	1500
ttggttgtat	ctaactttta	agggacagag	ctttacctgg	cagtataaaa	gatgggctgt	1560
ggagcttgga	aaaccacctc	tgttttccct	gctctataca	gcagcacata	ttatcataca	1620
gacagaaaat	ccagaatctt	ttcaaagccc	acatatggta	gcacaggytg	gcctgtgcat	1680
cggcaattct	catatctgtt	tttttcaaag	aataaaatca	aataaagagc	aggaaaaaaa	1740
aaaaaaaaaa	aaaaaactcg	ag				1762

<210> 51  
<211> 2059  
<212> DNA  
<213> Homo sapiens

<400> 51

cggcacgagg	tttggtctaa	gagttagcag	aactgtgttg	agatgggagc	acattttggtt	60
gtgttcaaca	cagaagcaga	gcagtgggaag	gaggcatgca	tccccagctg	ataccttcgg	120
ttattgctgt	agttttcatc	ttactttctc	gtgtctgttt	tattgcaagt	tgtttggtga	180
ctcatcacaa	cttttcacgc	tgtaagagag	gcacaggagt	gcacaagtta	gagcaccatg	240
caaagctcaa	atgcatcaaa	gagaaatcag	aactgaaaag	ttgctgaagg	gagcacctgg	300
aactgtttgtc	ctattgactg	gagaaccttc	cagttccaac	tgctattttc	ctccttactg	360
acaacaagac	gtgggctgag	agttgaaagg	aactgttcag	ggatgggggc	ccatctgatg	420
accatcagca	cgggaagctga	gcagaacttt	attattcagt	ttctggatag	acggctttcc	480
tattttccttg	gacttagaga	tgagaatgcc	aaaggctcgt	ggcgttgggt	gggaccagac	540
gccattttaac	ccacgccaga	gtattctggc	ataagaatga	acccgacaac	tytcaggggag	600
aaaactgtgt	tgtttctggc	ttataaccaa	gataaatggg	cctggaatga	tgttctctgt	660
aactttgaag	caagtaggat	ttgtaaaata	cctggaacaa	cattgaacta	gaaactcaga	720

aagtggctct	tgtgatggaa	agagaaaaga	aaaaccaatt	agaataaggc	agaatgtacg	780
tgcgtcattg	gaacacagaa	aacatgctgg	ttcatcacgc	gttttttagtc	ataatgggtct	840
tttttatttt	gtttgattca	ttcgagacaa	catgtgtgta	tgtgtgtgtg	tgtgtgtgta	900
gataatgtgg	tttttgtatg	gtgtttgatg	gaaggaataa	tctttctttg	ctttcttagt	960
agtatttcaa	ggtgtttact	tttcaattgg	tgtgcactga	atgcatgtat	ggaagaatag	1020
cgtagaataat	gcaatctctt	tgtcattttt	tcccctttct	cagactctta	gctcttaaaa	1080
ttcaaaagat	gggatattct	aactggtagt	ggtgcatcat	ttttaaccca	aatattgcaa	1140
gcacttttaa	gatttgaaac	cacattttta	ttgtttgatg	tttcattttc	agacttttta	1200
atgtcagtc	ttacaattac	attgcatgag	gaaaattttt	ccagaacaac	agtgtggaat	1260
agttctgaat	tatgctgttc	tacagatgga	aaaaagtcca	aatgccttta	aaaatttact	1320
tcttactcca	cccaacacgt	ttttgcaaag	caagaagtct	ttgtaagaca	ccttaaacia	1380
agtccttcaa	ttctacagca	gaggaaataa	aatccccag	aagccaaagg	gctcaccttc	1440
acattgttag	ttcatgacag	acccagggtg	gcttcattag	agataacata	cattcccttt	1500
gggtatcacag	gaagttactg	gggattactc	gacctatta	cttagctaac	gactggataa	1560
aattttcttaa	ttgtttgaag	taacattgta	ttcgtgtttg	cattattaat	ttgaatagaa	1620
aataatcaca	ttttcaaccc	atltatacaa	attgttaatg	tttctttaga	gctgtataac	1680
tatagtttga	actagcaagg	aagttattgt	tttgacaacc	agaaattatg	cttttctggt	1740
gcatgaaaca	ctaattgcaa	agggcagtc	catccaactt	taataaaata	tggtgggtctt	1800
tcttaaaatt	ttcaatttgc	taatttttcc	tggaaccta	actttcctac	tactactttt	1860
aaaagcagta	tttattttgt	aaattagcta	ctgatttttt	gggtttgaaa	tactgggtgtt	1920
tgattgtgtg	gttggtgtgg	tggtgatgga	tgctgtagc	tgtgttaaaa	gctgatactt	1980
ttagctgaaa	tggtgctaaa	taaatttgac	tgtgtttgac	atattcacta	aaaaaaaaaa	2040
aaaaaaacta	gagggggggg					2059

<210> 52  
 <211> 3282  
 <212> DNA  
 <213> Homo sapiens

<400> 52						
cccacgcgtc	cgacttaaaa	gagaagcttt	agctgccaaa	gattgggaaa	gggaaaggac	60
aaaaaagacc	cctgggctac	acggcgtagg	tgcagggttt	cctactgctg	ttcttttatg	120
ctgggagctg	tggtctgtaa	caactaggaa	ataacgtatg	cagcagctat	ggctgtcaga	180
gagttgtgct	tcccaagaca	aaggcaagtc	ctgtttcttt	ttcttttttg	gggagtgtcc	240
ttggcagggt	ctgggttttg	acgttattcg	gtgactgagg	aaacagagaa	aggatccttt	300
gtgggtcaatc	tggcāaagga	tctgggacta	gcagaggggg	agctggctgc	aaggggaaacc	360
aggggtgggtt	cggatgataa	caaacaatac	ctgctcctgg	attcacatac	cgggaatttg	420
ctcacaaatg	agaaactgga	ccgagagaag	ctgtgtggcc	ctaaagagcc	ctgtatgctg	480
tatttccaaa	ttttaatgga	tgatcccttt	cagatttacc	gggctgagct	gagagtcagg	540
gatataaatg	atcacgcgcc	agtatttcag	gacaaagaaa	cagtcttaaa	aatatcagaa	600
aatacagctg	aagggaacgc	atttagacta	gaaagagcac	aggatccaga	tggaggactt	660
aacggtatcc	aaaactacac	gatcagcccc	aactcttttt	tccatattaa	tattagtggc	720
ggtgatgaag	gcatgatata	tccagagcta	gtgttgga	aagcactgga	tggggaggag	780
cagggagagc	tcagcttaac	cctcacagcg	ctggatgggtg	ggtctccatc	caggtctggg	840
acctctactg	tacgcatcgt	tgtcttggac	gtcaatgaca	atgccccaca	gtttgcccag	900
gctctgtatg	agaccagggc	tccagaaaac	agccccattg	ggttccttat	tgtaaggta	960
tgggcagaag	atgtagactc	tggagtcaac	gcggaagtat	cctattcatt	ttttgatgcc	1020
tcagaaaata	ttcgaacaac	ctttcaaatc	aatccttttt	ctggggaaat	ctttctcaga	1080
gaattgcttg	attatgagtt	agtaaatctt	tacaaaataa	atatacaggc	aatggacggg	1140
ggaggccttt	ctgcaagatg	tagggtttta	gtggaagtat	tggacaccaa	tgacaatccc	1200
cctgaactga	tcgtatcatc	atlttccaac	tctgttgctg	agaattctcc	tgagacgccg	1260
ctggctgttt	ttaagattaa	tgacagagac	tctggagaaa	atggaaagat	ggtttgctac	1320
attcaagaga	atctgccatt	cctactaaaa	ccttctgtgg	agaattttta	catcctaatt	1380
acagaaggcg	cgctggacag	agagatcaga	gccgagtaca	acatcactat	caccgtcact	1440
gacttggggg	caccagggct	gaaaaccgag	cacaacataa	cctgtgctgt	ctccgacgtc	1500
aataacaacg	ccccgcctt	cacccaaac	tcctacaccc	tgctctgccc	cgagaacaac	1560
agccccgcgc	tgcacatcgg	cagcgtcagc	gccacagaca	gagactcggg	caccaacgcc	1620
caggtcacct	actcgtgtgt	gccgccccaa	gacccgcacc	tgccccctgc	ctccctgggtc	1680

tccatcaacg	cggacaacgg	ccacctgttc	gccctcaggt	cgctggacta	cgaggccctg	1740
caggcttttcg	agttccgcgt	gggcgccaca	gaccgcggct	ccccgcgct	gaacagcgag	1800
gcgctgggtg	cgctgctgg	tgctggacgc	caacgacaac	tcgcccttcg	tgctgtaccc	1860
gctgcagaac	ggctccgcgc	cctgcaccga	gctgggtgcc	cgggcgggcg	agccgggcta	1920
cctggtgacc	aaggtggtgg	cggtggacgg	cgactcgggc	cagaacgcct	ggctgtcgta	1980
ccagctgctc	aaggccacgg	agcccgggct	gttcgggtg	tgggcgcaca	atggggaggt	2040
gcgcaccgcc	aggctgctga	gcgagcgca	cgagccaaag	cacaggctgg	tggtgcttgt	2100
caaggacaat	ggcgagcctc	ctcgctcggc	caccgccacg	ctgcacgtgc	tcctgggtga	2160
cggctttctcc	cagccctacc	tgctctctcc	ggaggcggcc	ccggcccagg	cccaggccga	2220
cttgtctcacc	gtctacctgg	tggtggcggt	ggcctcgggt	tcttcgctct	tcctctctct	2280
ggtgtcctcg	ttcgtggcgg	tgcggtctgt	caggaggagc	agggcggcct	cggtgggtcg	2340
ctgtcccg	cccgagggtc	cttttccagg	gcctctggtg	gacgtgaggg	gcgctgagac	2400
cctgtcccag	agctaccagt	atgagggtgt	tctgacggga	ggccccggga	ccagtgaagt	2460
caagttcttg	aaaccagtta	tttcggatat	tcaggcacag	ggcctggga	ggaagggtga	2520
agaaaattcc	accttccgaa	atagctttgg	atttaatat	cagtaaagtc	tgtttttagt	2580
ttcatatact	tttgggtgtg	tacatagcca	tgtttctatt	agtttacttt	taaatctcaa	2640
atttaagtta	ttatgcaact	tcaagcatta	ttttcaagta	gtataccct	gtgggttttac	2700
aatgtttcat	catttttttg	cattaataac	aactgggttt	aatttaatga	gtattttttt	2760
ctaaatgata	gtgttaaggt	tttaattctt	tccaactgcc	caaggaatta	attactatta	2820
tatctcatta	cagaaatctg	agggtttgat	tcatttcaga	gcttgcatct	catgattcta	2880
atcactttctg	tctatagtgt	acttgctcta	tttaagaagg	catatctaca	tttccaaact	2940
cattctaaaca	ttctatatat	tcgtgtttga	aaaccatgtc	atttatttct	acatcatgta	3000
tttaaaaaaga	aatattttctc	tactactatg	ctcatgacaa	aatgaaacaa	agcatattgt	3060
gagcaatact	gaacatcaat	aataccctta	gtttatatac	ttattatttt	atctttaagc	3120
atgctacttt	tacttgccca	atattttctt	atgttaactt	ttgctgatgt	ataaaacaga	3180
ctatgcctta	taattgaaat	aaaattataa	tctgcctgaa	aatgaataaa	aataaaacat	3240
tttgaaattt	gtgaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aa		3282

&lt;210&gt; 53

&lt;211&gt; 1860

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 53

aattcggcac	gagcagcagg	gagaagagga	tgatgagcat	gccaggcccc	tggccgagtc	60
cctgtcctcg	gccattgctg	acctgctctt	ctgcccggac	ttcacgggtc	agagccaccg	120
gaggagcaact	gtggactcgg	cagaggacgt	ccactccctg	gacagctgtg	aatacatctg	180
ggaggctgggt	gtgggcttcg	ctcactcccc	ccagcctaac	tacatccacg	atatgaaccg	240
gatggagctg	ctgaaactgc	tgctgacatg	cttctccgag	gccatgtacc	tgccccccagc	300
tccggaaagt	ggcagcacca	acccatgggt	tcagttcttt	tgttccacgg	agaacagaca	360
tgccctgccc	ctcttcacct	ccctcctcaa	caccgtgtgt	gcctatgacc	ctgtgggcta	420
cgggatcccc	tacaaccacc	tgctcttctc	tgactaccgg	gaaccctctg	tggaggaggc	480
tgcccagggtg	ctcattgtca	ctttggacca	cgacagtgcc	agcagtgccca	gccccactgt	540
ggacggcacc	accactggca	ccgccatgga	tgatgccgat	cctccaggcc	ctgagaacct	600
gtttgtgaac	tacctgtccc	gcctcctatg	tgaggaggac	ttccagttca	tcctcaaggg	660
tatagcccgg	ctgctgtcca	acccctgtct	ccagacctac	ctgcctaact	ccaccaagaa	720
agatccagtt	ccaccaggag	ctgctagtct	tcttctggaa	gctctgcgac	ttcaacaaga	780
aattcctctt	cttcgtgctg	aagagcagcg	acgtcctaga	catccttgtc	cccactctct	840
tcttccctcaa	cgatgcccg	gccgatcagt	gtaagaccag	gggtgggggtg	ggatgctggg	900
ggctttcctg	gcggcgaggt	ggagggtctg	ctatctgccc	tgactcccag	gagctcagcc	960
tggcactctg	acctcctagg	agggaggcct	ccaaaatggt	cactgcctgg	gccccctcctt	1020
ccccctccca	cccaggtgac	ccctcccacg	ccacctgccg	cagctcgggt	gggcctgatg	1080
cacattgggtg	tcttcatctt	gctgcttctg	agcggggagc	ggaacttcgg	gggtgcggctg	1140
aacaaaacct	actcaatccg	cgtgccatg	gacatcccag	tcttcacagg	gacccacgcc	1200
gacctgctca	ttgtgggtgtt	ccacaagatc	atcaccagcg	ggcaccagcg	gttgacgccc	1260
ctcttcgact	gcctgctcac	catcggtggtc	aacggtaggg	gcccggagct	tgcaactccc	1320
gcgctcacca	ccagggtgggtg	tcagccacag	gccaacctgc	cacttgctcc	aggtttcaag	1380
ccccctccagc	ccacagaaag	aacgcaagct	tgtgtaccct	tcatagtagc	cacattaaaa	1440



aaaaaggggtg	acaacaattg	aataatatgt	tttatctatc	atcatctcaa	catgtagcca	1500
atatgaaaat	tactaatgag	atattttaata	ttcccttctc	acactcagtc	ttgtcgccca	1560
ggctggagtg	cagtggcacc	atctcagctc	actgcaacct	ccgccccttg	ggttcaagca	1620
attctcatgc	ctcagcctcc	tgagtaggtg	ggattacagg	cgtgtgccac	cacacctggc	1680
taatttttgt	atcttttagta	gagacggggt	tttgccatgt	tgtccagggt	ggctctcgaa	1740
tcctgacctc	agatgatatg	cccacctcag	cctcccaaag	tgctgggatt	acaggcatga	1800
gccactgcat	ctggcctttt	taaaaaaaaa	aaaaaaaaaa	actcgagggg	gggcccggta	1860

&lt;210&gt; 54

&lt;211&gt; 770

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 54

aattcggcac	gagccgaggg	gcgaatccat	acatccgtcg	tgtggactcg	ccctgacatc	60
ccctcaacgc	ttagactcgg	cattcttttc	attgcctgat	gccgtttcat	gccttctttg	120
cgtttcctcg	ctctggccct	cctgctggcc	atccttcccg	cacttcccaa	cgcccatgcc	180
gcgcggggga	ttggcgccct	gatcggcggc	ggctcccagg	cctcagccaa	ggaagagccc	240
cagagcaacg	cgcaaccag	cgccgatgag	cgcaacaac	gactgctcag	ccaggccgag	300
gaaacccggc	agcgccctgac	cgatctcaag	gccgaactcg	ccggcgccacc	caaggagatc	360
agcgaggccc	agcgtacgct	ctccaaactg	gtgagcgagg	acaacagcga	tctgcccagag	420
cgccctctcca	agctttcggg	gccggtagctg	gagcaacgcc	tggcgggccc	cgtaggacgag	480
ctcgccctct	ggcaacaagc	gtcagcgcg	gccaacagca	tgctcatcag	cgcgagaccc	540
cgccccgagc	gcgcccaggc	cgatatcagc	aagaaccagt	tgcgcaccca	cgagatcaac	600
ggcctgctga	aaagcgggtcg	ggagaacaac	aagccgctga	cgagcgaacg	tcgcgcgctg	660
ctcgagagta	cttctagagc	ggcgcggggc	ccatcgattt	tccaccggg	tgggggtacca	720
ggtaagtgtg	cccaattcgc	cctatagtga	gtcgtattac	aattcactgg		770

&lt;210&gt; 55

&lt;211&gt; 1093

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 55

cagattcggc	acgaggtttt	ttttaagatt	tattttttta	aaattatatt	tgtggacttg	60
ggtatcaatg	atggcaccta	cttttgggaa	tctgtagctg	tgctttgaga	attgccatcg	120
gtcatgtgtt	gcaccgttct	ctgtatgttt	acgtcccttg	gactggcttc	tcccaggatt	180
cttttctgtt	tttgtttttt	tgatttgggc	tttatttttt	tctgtgtact	gtactatatt	240
gtaaaagggg	tttttagcaga	gacttttagt	tttggggcaa	gaggagaaca	ggaatgctgg	300
gctgtttact	ttaggtggag	aacctatctt	cagacctttg	gactattttc	tttcaactgc	360
agtgtataga	aaaaccacaa	tacgacctca	gagcagagta	ttaatgaaaa	gcacaaaaaa	420
aggaactaag	ttcagcgagg	ggtgggggga	ggggggagat	ttttcttttg	aaaaataatg	480
actcttagga	catttgtttt	tcagttcaag	tgctcttcag	cactgtcttg	tctcccaata	540
taccaacca	ctggcacatt	tttctcttgt	tttctctctc	cgatttttgt	ctgtctcttc	600
agttaagtgt	ttcttctctt	tgtgcccccc	gctggtgacc	ctctgcttcc	ctctctcttt	660
ccctttggca	gctgcaatac	acagtgttat	tttggggaaa	taaatctagc	aaagcctcgc	720
cttccatgcc	gagcgtcttc	ttggctctga	gagggaaagg	tctgtccttg	ggatgctctc	780
tgggtctttt	tccccctaag	tctttctctt	tcccatcata	cccttccctg	cccaccttgt	840
tttctgttct	ccttttatta	ggaattccca	agtgaatttt	attaatgtgg	gagtgggaaca	900
gatgctaaaa	gctatccagg	attttgtttc	tgtttgtttt	aaattttgtg	gttccttccc	960
tttctctccc	ctcccatgcg	taagacgttc	tgtgtaacct	ccattaaatt	tggtacaaaa	1020
ccactcgcca	gagctgtggt	gtcagaaaaa	taaaatatat	tgtttcttam	aaaaaaaaaa	1080
aaaaaaaaact	cga					1093

&lt;210&gt; 56

&lt;211&gt; 632

<212> DNA  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (29)  
<223> n equals a,t,g, or c

<220>  
<221> SITE  
<222> (46)  
<223> n equals a,t,g, or c

<220>  
<221> SITE  
<222> (94)  
<223> n equals a,t,g, or c

<220>  
<221> SITE  
<222> (162)  
<223> n equals a,t,g, or c

<400> 56  
cgcaattaat gtgagtttagc tcactcatna ggcaccccag gctttncact ttatgctttc 60  
cggctcglat gttgtgtgga attgtgacgg atancaattt cacacaggaa acagctatgc 120  
catgattacg ccaagctcga aattaaccct cactaaaggg ancaaaaagct ggagctccac 180  
cgcggtggcg gccgctctag aactagtggg tccccggggc tgcaggaatt cggcacgaga 240  
ctatgtatat atgtttaata tctgtctttt gaaatgcaga aatagtttta atgtttcttt 300  
gtctattttt cttttttttt aatgctaccc agggaaatat ttccatatca tttttaagtg 360  
gcctgcctca atgtatatatt atttcttttg aaacaaaaag gttctggaaa ctgtttttct 420  
gtagctttta atgaataggt gagcaaaatc tatatgggat gtaatttttt tgttcagtct 480  
cttaaaaaat actttgtttt ggtacatttg gttgtgcttg tggggaaaat aaaaacgcag 540  
agatccttat atatttatgt taaagtaata ttttattatc tacataaaac agaaatgcac 600  
aataaaaaaa aaaaaaaaaa gctcgagggg gg 632

<210> 57  
<211> 2687  
<212> DNA  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (1614)  
<223> n equals a,t,g, or c

<400> 57  
gtacaccatg ggcctccacc tccgcccta ccgtgtgggg ctgctcccgg atggcctcct 60  
gttctctttg ctgctgctaa tgctgctcgc ggacccagcg ctcccggccg gacgtcacc 120  
cccagtgggt ctggtccctg gtgatttggg taaccaactg gaagccaagc tggacaagcc 180  
gacagtgggt cactacctct gctccaagaa gaccgaaagc tacttcacaa tctggctgaa 240  
cctggaaactg ctgctgcctg tgcattcatt actgctggat tgacaatatc aggctggttt 300  
acaacaaaac atccagggcc acccagtttc ctgatgggtg ggaatgtacgt gtccctggct 360  
ttgggaagac cttctcactg gagttcctgg accccagcaa aagcagcgtg ggttcctatt 420  
tccacaccat ggtggagagc cttgtgggct ggggctacac acggggtgag gatgtccgag 480  
gggctccccta tgactggcgc cgagcccaa atgaaaacgg gccctacttc ctggccctcc 540  
gcgagatgat cgaggagatg taccagctgt atggggggcc cgtgggtgctg gttgccca 600  
gtatgggcaa catgtacacg ctctactttc tgcagcggca gccgcaggcc tgggaargaca 660

agtatatccg	ggccttcctg	tactgggtg	cgccctggg	gggctggcc	aagaccctgc	720
gcgtcctggc	ttcaggagac	aacaaccgga	tcccagtcac	cgggcccctg	aagatccggg	780
agcagcagcg	gtcagctgtc	tccaccagct	ggctgctgcc	ctacaactac	acatgggtcac	840
ctgagaaggt	gttcgtgcag	acaccacaa	tcaactacac	actgcgggac	taccgcaagt	900
tcttccagga	catcggttt	gaagatggct	ggctcatgcg	gcaggacaca	gaagggctgg	960
tgggaagccac	gatgccacct	ggcgtgcagc	tgcactgcct	ctatggtact	ggcgtcccca	1020
caccagactc	cttctactat	gagagcttcc	ctgaccgtga	ccctaaaatc	tgctttggtg	1080
acggcgatgg	tactgtgaac	ttgaagagtg	ccctgcagtg	ccaggcctgg	cagagccgcc	1140
aggagcacca	agtgttgctg	caggagctgc	caggcagcga	gcacatcgag	atgctggcca	1200
acgccaccac	cctggcctat	ctgaaacgtg	tgtccttgg	gccctgactc	ctgtgccaca	1260
ggactcctgt	ggctcggccg	tggacctgt	gttggcctct	ggggctgtca	tggccacgc	1320
gttttgcaaa	gtttgtgact	caccattcaa	ggccccgagt	cttggactgt	gaagcatctg	1380
ccatggggaa	gtgctgtttg	ttatccttcc	tctgtggcag	tgaagaagga	agaaatgaga	1440
gtctagactc	aagggacact	ggatggcaag	aatgctgctg	atggtggaac	tgctgttracc	1500
ttaggactgg	ctccacaggg	tggactggct	gggccctggt	cccagtcctc	gcctggggcc	1560
atgtgtcccc	cctattcctg	tgggcttttc	atacttgctc	actgggccc	ggcncsgcag	1620
ccttccctatg	agggatgtta	ctgggctgtg	gtcctgtacc	cagaggtccc	agggatcggc	1680
tcttggcccc	tgggtgacc	cttcccacac	accagccaca	gataggcctg	ccactggtca	1740
tgggtagcta	gagctgctgg	cttccctgtg	gcttagctgg	tggccagcct	gactggcttc	1800
ctgggcgagc	ctagttagctc	ctgcaggcag	gggcagtttg	ttgcgttctt	cgtgggtccc	1860
aggccctggg	acatctcact	ccactcctac	ctcccttacc	accaggagca	ttcaagctct	1920
ggattgggca	gcagatgtgc	ccccagtcct	gcagctgtgt	tccagggggc	ctgatttcc	1980
cggatgtgct	attggcccca	ggactgaagc	tgcttccctt	caccctggga	ctgtgggttc	2040
aaggatgaga	gcaggggttg	gagccatggc	cttctgggaa	cctatggaga	aagggaatcc	2100
aaggaagcag	ccaaggctgc	tgcagcttc	cctgagctgc	acctcttgct	aacccacca	2160
tcacactgc	accctgccct	agggtctcac	tagtaccag	tgggtcagca	cagggctgag	2220
gatggggctc	ctatccaccc	tggccagcac	ccagcttagt	gctgggacta	gcccagaaac	2280
ttgaatggga	ccctgagaga	gccaggggtc	ccctgaggcc	cccctagggg	ctttctgtct	2340
gccccagggt	gctccatgga	tctccctgtg	gcagcaggca	tggagagtca	gggctgcctt	2400
catggcagta	ggctcctaag	gggtgactgg	ccacaggccg	agaaaagggt	acagcctcta	2460
ggtgggggtc	ccaaagacgc	cttcasgctg	gactgagctg	ctctcccaca	gggtttctgt	2520
gcagctggat	tttctctgtt	gcatacatgc	ctggcatctg	tctccccttg	ttcctgagtg	2580
gccccacatg	gggctctgag	caggctgtat	ctggattctg	gcaataaaag	tactctggat	2640
gctgtaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	ggcgccc		2687

<210> 58  
 <211> 619  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (526)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (619)  
 <223> n equals a,t,g, or c

<400> 58						
tcatcagctt	tcacctggat	tatcgtgaca	gcctcctact	gcttctctat	catgtggcca	60
gagctatctt	cctaaaaatgc	attgcatagt	tgatcaagtc	actctctggc	ctaaaaacctt	120
ccttggctcc	ctgctgccct	caggataaag	tctggacccc	tcagcatggc	ttgtgagact	180
catgggtgtc	ttgtccctgc	tcacctctct	ggctctcatc	cttgccttct	tgcattctgg	240
gtcccagcct	cctgtatcca	gagatgcagt	ggctctccat	tgccactctg	attcctcctt	300
tcttttggtc	acagagaaag	ggtactttct	ctgtcaaate	tcaacttaga	cttgacttcc	360
tccaaggagc	tttggctata	ctctctcttc	ccgaccccca	ccctggcata	ctacacagat	420

cactctgggc	tcacttgcc	gcctaattggt	catctcccca	gtagactgta	agctccttga	480
gggcaaggat	tgtgttgga	tttttgtatt	aacagtgcct	ggcttngtgc	ctggcaccta	540
gaaagcactc	aataaatggt	tgtttaatga	aaaaaaaaa	aaaaaaaaa	aaaaaaaaa	600
aaaaaaaaa	aaaaaaaaa					619

<210> 59  
 <211> 1378  
 <212> DNA  
 <213> Homo sapiens

<400> 59						
tgcacttctg	ggatgtagat	tctcctttat	tgaattcttc	atcttcactc	gcagtatgat	60
ttattttttt	cacactgtat	tttactactgc	atatgtttct	cttataacga	gggtcacttt	120
cttctgccc	aggattggaa	tcaattacag	ctatgcaatg	ttttttgttt	agcatcttct	180
taattacagg	cttggcagaa	gagttttgtg	aacaactctc	catttcactg	gctgaagaag	240
aaatacaact	gtcctctaca	gttgaacatt	tctgtatgac	agcattttca	tggtaatgag	300
gattagggtc	atactttggt	ttttccaaac	cggggaaaca	aacaacagcc	aaaaaccagt	360
gtgcagcttc	attaaggggt	acaaaaataa	aatccttctc	aaaaatatct	acgtgccggg	420
tccatgtttt	tactctccca	tgccgttttt	gctgtattga	cagattagtt	gtttcatgat	480
ttctcctctc	tctctgatta	aggcgtttat	agaaaaaaga	actgaatata	tgaattcggg	540
cagcgtcttc	cttcttcagt	ttttcaagca	ccaagtattt	caaataaaaag	tctataataa	600
catcatttaa	aaattctcct	tcatttagac	agtgcaggtc	ctcattggta	acagagatgc	660
ctcccttagc	tggagggtgt	ggatatacta	tcaacttttc	tactgggcca	atgaagatgg	720
tgtgtgtttt	tccagtttct	tcttcttcac	caaaaaactg	aaattcttgt	ttgcttctaa	780
gttgtatttt	agattcaaat	gatacagttt	taattttgtt	ttccttttgc	ccacaacttc	840
ctttgatgct	ctcttcacag	gttcttgtac	aggcaacaag	tctgccatta	gcttcttcaa	900
agggaatttt	cgcaaaaaaa	ttggagatgt	tattctttat	accaatttca	ttaatgatac	960
tttcaaatat	catatttgc	ggaggatcaa	ggccattttg	aaaaattaaa	attatatatt	1020
gttcttctaa	atttggtaat	ttatttactc	ctttacaatc	attccaaact	ttatcctcct	1080
tattcatttg	cagttggatg	ctcagctttt	gataaactgc	tggaattgct	tgaagaaaca	1140
ctacaggtaa	ttttcggaca	ttacaccatt	cacatttagt	tagatcagag	gtatttaata	1200
taatctctac	aggatcatgg	tctgtgtcgt	ctagctgtat	cttgataaaa	tctaaacaaa	1260
aaattacagg	ctctattaac	agccggaaga	gtgttccctac	tgcgtcgacgc	ggccgcgaat	1320
tcccgggtcg	acgagctcac	tagtcggcgg	cgcctctaga	ggatccctcg	aggggccc	1378

<210> 60  
 <211> 1126  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (21)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (35)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (49)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE

&lt;222&gt; (99)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (1012)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 60

tccggcttcg	taatgtggtg	ngaaattgta	acggnataac	aatttcacnc	aggaaccagc	60
tatgcccattg	attacgccaa	gttcgaaatt	aaccctcant	aaaggaccaa	aagtgaagtt	120
ccaccgcggt	gccggcccg	tctagaaact	agtggatccc	ccgggctgca	ggaattcggc	180
acgagccggc	ccgtaccgcc	aggcgatcgc	gctgatggcg	gcgctggcag	cagcggccaa	240
gaaggtgtgg	agcgcgcggc	ggctgctggt	gctgctgttc	acgccgctcg	cgctgctgcc	300
ggtggtcttc	gccctcccg	ccaaggaagg	ccgctgcttg	tttgtcatcc	tgctcatggc	360
ggtgtactgg	tgcacggagg	ccctgccgct	ctcagtgcag	gcgctgctgc	ccatcgtcct	420
gttcccccttc	atgggcatct	tgccctccaa	caaggtctgc	ccccagtact	tcctcgacac	480
caacttcctc	ttcctcagtg	ggctgatcat	ggccagcgcc	attgaggagt	ggaacctgca	540
ccggcgaatc	gccctcaaga	tcctgatgct	tggtggagtc	cagccggcca	ggctcatcct	600
ggggatgatg	gtgaccacct	cgctctgttc	catgtggctg	agcaacaccg	cctccactgc	660
catgatgctt	cccattgcca	atgccatcct	gaaaagtctc	tttgccaga	aggaggttcg	720
aaaggacccc	agccaggaga	gtgaagagaa	cacaggaata	gaacccaata	cttctctctc	780
tgaggaaagg	ctgaaacttc	aagctcccc	tgtgataaga	cttggtcaga	taactgagtc	840
tggtcaatgg	aatatgagtg	gaaatgatgt	gtgcaacttc	cgggttctgt	ccttctctgc	900
gggtggaatg	tgaatatgat	ggcacctggg	acccaaagac	aggagccaca	tcttgagaga	960
tagatggcag	atctgcccc	gtggcttttg	atcatttacc	tcagtgaaca	cnacaagcat	1020
tatccatgaa	accatagggt	ttgtgtgcta	gttctagttt	ttaaaatatg	aattaaatta	1080
aatacgtatc	tgttaaaact	taaaaaaaaa	aaaaaaaaaa	ctcgag		1126

&lt;210&gt; 61

&lt;211&gt; 2078

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (337)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (492)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 61

gattattttc	aatagattgt	caataaatta	tcaaattatg	gaaatatact	ttagattcct	60
gtgataatgc	tttgtgcatg	ttttctatta	tttagaaaat	gaaaacagca	gtcatttgac	120
cagccatggt	atgacaccag	ataatattct	gcaaggatct	tccaaacagg	gagaaactgc	180
aatgggaaca	atctttggat	atctacattg	tgtcaaattg	tatgtgcttt	atttcatttt	240
catacttata	acagctgtat	atcacagttt	ttattatcca	cattatagag	gcaaggcact	300
gatatcagga	acttgagtaa	ttggcttaca	cagcaantgt	agttgaaata	gaagccagga	360
ctgtctgatt	cccaagacag	tggtcttgta	ctgcaggtaa	gcctgggtgt	gctgggkttc	420
cctagcactg	agtattgcag	aaatacggag	atttattttac	tcagggtgaat	gatgaagcag	480
aatcttttatt	tngcttttta	aattatkgat	gggkckkgag	ttaagctagt	tcaccaatga	540
gtctcaaagg	gtgaagaact	ggcctgtgta	gttttaaaac	aaagcagcag	ttttcttcac	600
ttgtagatat	tatatttaat	ttaccttgca	tttccaatag	cattaaatat	tagaagcgta	660
taagcttttt	ctttttatgc	ctaggggata	tggcttaaca	ctgagaaaca	cttgagaaac	720
actgccacat	tctaaggaac	aaccaataat	aagaaaaatt	atttcatata	aaaggatgca	780

gcttgcatat	aaaatgctga	attcaactga	agttcattat	gctgccttta	tgtgagctca	840
ggggtgacct	tcatgctctg	gactcttaaa	ggatccagtg	tcggatacag	ctctgccatt	900
ggagcaatgt	gttccataac	taggggtgcc	taagtatatc	ccttcttctg	ktatctatag	960
tttttcttta	aaggtgtgct	tttcaaaaaca	tgatgttatt	tcatatacmt	atatatatat	1020
atgcacacat	acatatacmt	catgcatgka	tcatatatgt	gtgtgtatgt	atgtatctat	1080
cttcaatctt	tcaggaaagt	ttgtaaaawa	tagccctaag	tagaattatg	gggaaacagt	1140
tgcagaactg	ctgaaattgg	ctcagatgaa	aaggaaagag	caaaaatcta	ggttattcac	1200
aggaagttag	tccagctgtg	ttttgcgagg	catgaaccag	gctgaatcag	agcaagaaca	1260
ggttgcaatc	ttaggtcaca	tggcaattaa	attccagcta	tagaatattg	aagacatgta	1320
agtgcctaat	ttaaggatta	ttcaatgcag	tgaataaatt	tatatcattc	ccaaaagtat	1380
gcacatcaga	taatgagtta	actgcatgta	acttgaaaac	cttaaatatt	taactgaaag	1440
gattttgtgc	atagaaagat	gtcttagtca	atgtaaaaat	gccagtgatt	tgtgtagcat	1500
gtcctttgct	aaaagtcatt	atataaaactg	aagtcactta	tctttgtttc	caatgtgata	1560
atattccaca	tttaacaggt	gataattatt	gaaaatatta	gtttgcctca	ctaattattgt	1620
gtccgagata	ttctagttta	aaccaagtaa	ctaatactgt	aaaagaagta	ttccttttat	1680
gaagccagca	ttatcctgat	actagaactg	ggaagagaca	cacacaaaaa	gaaaacttca	1740
ggccaatatc	gctgatgaac	atcaatgtga	aaatcttcaa	taaaatactg	gcaaaccgaa	1800
tccagcagca	tgaggcgggt	ggatcacaa	gtcaggagat	cgagaccatc	ctggctaaca	1860
tggtgaaacc	ccatctctac	taaaaataca	aaaaattagc	cgggcatggg	gatgggcacc	1920
tgtagtccca	gctactgggg	aggctgaggc	aggagaatgg	cgtgaacctg	ggaggcggaa	1980
cttacagtga	gccgagattg	caccactgca	ctccagcctg	ggtgacagag	caagactctg	2040
tctcaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aactcgta			2078

<210> 62  
 <211> 762  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (10)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (12)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (42)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (219)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (747)  
 <223> n equals a,t,g, or c

<400> 62	
cgagtcagtn	ancgagaaag
tgcccgatcc	attaatgcag
cgcaacgcaa	ttaatgtgag
cttccggctc	gtatgttgtg
cggaagagcg	cccaatcgca
ctggcacgac	aggtttcccg
actgaaaagc	ggccagtgag
cccaggcttt	acactttatg
aatttcacac	aggaaacagc
	60
	120
	180
	240

tatgaccatg	attacgccaa	gctcgaaatt	aaccctcact	aaagggamca	aaagctggag	300
ctccaccgcg	gtggcgccg	ctctagaact	agtggatccc	ccgggctgca	ggaattcggc	360
acgagcacaa	atgcctatac	aaaacttttt	tccatgtcct	caatactttg	tatcagatta	420
tttatatttc	tcatggtttg	gtttcttttt	cttgtcttca	ttttctgaa	ggtaaaagg	480
gattttttcc	ccccttttct	gatctgtaac	ttgttttgta	tttggatgat	tacaggcgtg	540
agccaccgcc	tccagccaca	aatacttttt	agtcgccata	agcataatca	agaaatcatt	600
ttgcaaatgg	taagcttttc	atgtttgtga	ttctttccca	tgattagaga	ggtaaaaagt	660
gsccctgggat	gcataaagat	gtcgtagacm	tmtgcaatct	gccacattat	ttctttaaaa	720
aaaaaaaaa	aaaactcgag	gggggggnacg	gaaccagggg	cg		762

&lt;210&gt; 63

&lt;211&gt; 1094

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 63

tcggcacgag	gtcaatcaag	tgaaatatca	tgtaaactgt	ccagcagctt	tgaaagtaga	60
gaatgaacaa	ggcccccttc	ccacccaccc	tgtggaaagc	ccgtctgggt	tggtgtcctc	120
ctggacagcg	tcttgccggt	cacctttggc	catctcccgg	tgcgtgggtc	agatgtgggt	180
cctgcttttc	tgccccctcc	ctcctctgtg	cctgectgcc	tctgctgtgc	cggggccagt	240
ccttggtggc	cagtggagt	gacaccagct	gcgactgcgt	gggaggggct	ggcattgccg	300
ctgccactgc	agggcttggg	cggctgacat	gggacgaggg	ttgcacagct	gccagctcct	360
gtctcgctga	ctttttttat	acagttttgt	ctggggccacc	gccttcagt	ccacggggccc	420
cttgccgttc	aggctgtctc	tcatagatga	acaaggccct	gccccgtgtc	cttacccttt	480
agagctgttt	aaattcaaat	gaactgaaac	tgaatatgaa	aaatccaggc	cctcagccgc	540
ccaggccatg	tttcaagtgc	tccatggcca	catgtggctg	gtggacagt	cagctctaga	600
acattccatc	accacagagg	gttctgctgg	acagtggcct	tgggggctgt	tttgagggtc	660
cgcctgtcag	tctcctggca	tcaaagtcac	tctgccattg	tcaagttaca	gttattttcc	720
ttttacctcc	aagccactat	gtgcgtatgc	tgctatgtgt	ctgtatttcc	cgctaaactt	780
cctgtcacgg	aggggaagggt	gccacaggcc	cagctcctgg	aggggggttg	gatgtctggg	840
tggggggaag	ggtgccacag	gcccagctcc	cggagggggg	ttggatgtct	gggtgggggc	900
cagggcgcca	cagttccagc	tcccggaggg	ggtttggatg	tctgggcggg	gctttcttca	960
tgttccatgt	atgataacgg	tgactggggg	gtttacagag	agaggcatac	aaatagtgtt	1020
ggagtgttgt	tttcgcta	ataaatgttt	gaatagctaa	aaaaaaaaaa	aaaaaactcg	1080
agggggggcc	cggt					1094

&lt;210&gt; 64

&lt;211&gt; 1361

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 64

cccgggtcga	cccacgcgtc	cgaatttctt	tagtgtat	tatgtact	tgtattaatt	60
acattgggta	gactgttcca	taaatgttga	atagatccta	tttgttgggt	aatgctattg	120
ttgtgtatcc	tgtgattttt	ctgtgtagtt	ggctctatcag	ttgttgggag	aagagtgttg	180
aagtctacaa	ctataattgt	gtatttgtct	attactcctt	tcagttcttt	cagttccatt	240
tcacatattt	tccagctctt	gattgggtgca	cattaatgat	tactatgtct	tcttggtgaa	300
tagaccattt	tataatttta	tgatgtacct	ctcttgtcac	tttctttgct	ctgaagtcta	360
cctgatatta	atatagtcac	tctagctttc	tttggagtaa	tgttagtgtg	atatactttt	420
tttcgtcctt	ttacttttaa	ccctatctat	atcattatat	tattttaagt	gagttttttg	480
tagatagcac	ttactttgtg	ctgtcatggt	tttaaatcca	ctgtacaagt	acaaatctgt	540
ctttattttt	cattttttatc	cttatttttga	gacagagtct	cactttgaca	ctcagcctag	600
aatgcactgg	tgtgatctcg	gctcactgca	acctctgcct	ccgaggggtca	agcaatcctt	660
tcacctcagc	ctcccaagta	gctgggacta	caggcatgaa	ccaccatgcc	tggctaattt	720
tttttttttt	tttttttgta	gagatgggga	tttaccatgt	tgctgagggt	ggccctgaac	780
ccctgggctc	aaacgacctg	cccagctcag	cctcccagag	aaatatgatt	ttaattagca	840
catttagatc	atttactttc	aatggaatta	ttgatgtatt	acaattcaag	tctgtcattt	900

tattttattgc	cttccatttt	tcctccggtt	tttcatttct	gtgctttttt	tctttgcatg	960
taaacaatggt	ttaggattct	gttttgatgt	atttagagta	tttttttagtg	tatcggttta	1020
tgtagctttt	ctagtggatc	ctttaagtat	tacattatat	atacataact	aagaagtata	1080
tcagtgtcat	cattttacca	gttcaagtaa	agtacagaaa	tctttgtccc	ttttacctgt	1140
cctgcttata	atgtaaagt	cttaaagatt	tccttgacac	acatttagaa	ccatatcaca	1200
gtgttgtttt	tgcttcagtt	gtgaaatata	atttagaaac	tcagtagaat	aaaaacccat	1260
tatatatttt	tataaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1320
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	agggcgggccg	c		1361

&lt;210&gt; 65

&lt;211&gt; 947

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (67)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 65

ctttctatag	ttaaagccgg	tacgcctgca	ggtaccggtc	cggaattccc	gggtcgaccc	60
acgcgtnccg	tttctgggtt	ctaggccctg	cttccctgce	cctttgctgc	agaagggcag	120
ctgaaggctc	accctagaaa	ccgggcctgg	tgggtcttac	ccggctcact	ccctcccttg	180
tccttacaca	tacaggaaga	caagacctga	gtggtgctgt	cctttgtgtcc	gtcgtgtatg	240
gtctccctg	tcttcatttc	ttctcactct	gtctctaaac	ctctctctct	ctcccttccc	300
cctcagta	tartctacag	acctatgtgc	gtgtycctat	ccttctgtcc	ttttctctct	360
tcagctctcc	ctgcctctca	cacacaattt	tacatgcccc	gaggagccaa	gtttgggaca	420
tttaccctcc	aggcatctgt	gtccccctct	gaagagaaaa	cacacagctt	cacacatcca	480
ggcatagggg	gcaagctctt	ggggcatcag	gaccctggag	caccaggtcc	ttcctggaat	540
attagatcca	cctggagcac	caggtctctc	taagtctcac	ctgggggaatt	cggtcccacc	600
tgggtcacca	gttcccacct	agagcactgt	gtcctgcccc	agagcacaaa	gacctgctcc	660
tcccagagact	ctctctgact	gcagccaggc	atagtaccct	tgcctgtgtt	tgctccctgg	720
tccacagatt	tgggtggctg	gcaggtgctt	ggacagtgat	gaggtcttgc	cgccttaact	780
gtccccccca	gtcacttctc	ccacaggccc	agcaggacgc	agtcttgagg	atcagggtatt	840
ctacagctgc	attaaaaatca	atcctatcca	aaaaaaaaaa	aaaaagggcg	gccgctctag	900
aggatccaag	cttacgtacg	cgtgcatgcg	acatcaagct	ctgaaga		947

&lt;210&gt; 66

&lt;211&gt; 1376

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (18)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 66

ggcggggcgg	gcggaagngg	cggckgcgcg	gccggggcag	ccatgtcgcc	attgtctgcg	60
gcgcggggcg	ccctgcgggt	ctacgcggta	ggcgccggcg	tgatcctggc	gcagctgctg	120
cggcgctgcc	gcgggggctt	cctggagcca	gttytcccc	cacgacctga	ccgtgtcgct	180
atagtgcagg	gagggacaga	tggcattggc	tattctacag	cgaagcatct	ggcgagactt	240
ggcatgcatg	ttatcatagc	tggaaataat	gacagcaaa	ccaaacaagt	tgtaagcaaa	300
ataaaagaag	aaaccttgaa	cgacaaagtg	gaatttttat	actgtgactt	ggcttccatg	360
acttccatcc	ggcagtttgt	gcagaagtgc	aagatgaaga	agattcctct	ccatgtcctg	420
atcaacaatg	ctgggggtgat	gatgggtccc	cagaggaaaa	ccagagatgg	attcgaagaa	480
catttcggcc	tgaactacct	agggcacttc	ctgctgacca	accttctctt	ggatacgctg	540



aaagagtctg	ggccccctgg	ccacagtgcg	aggggtggta	ccgtctcttc	tgccacccat	600
tacgtcgctg	agctgaacat	ggatgacctt	cagagcagtg	cctgctactc	accccacgca	660
gcctacgccc	agagcaagct	ggccccctgtc	ctgttcacct	accacctcca	gcggctgctg	720
gcggctgagg	gaagccacgt	gaccgccaac	gtgggtggacc	ccgggggtggg	caacacggac	780
stctacaagc	acgtgttctg	ggccacccgt	ctggcggaaga	agcttctcgg	ctgggttgctt	840
ttcaaagacc	ccgatgaagg	agcgtggact	tccatctacg	cagcagtcac	cccagagctg	900
gaaggagtgtg	gtggccgtta	cctatacaac	gagaaagaga	ccaagtccct	ccacgtcacc	960
tacaaccaga	aactgcagca	gcagctgtgg	tctaagagtt	gtgagatgac	tgggggtcctt	1020
gatgtgaccc	tgtgatatcc	tgtctcagga	tagctgctgc	cccaagaaac	acattgcacc	1080
tgccaatagc	ttgtgggtct	gtgaagactg	cggtgtttga	gtttctcaca	cccacctsc	1140
cacagggtc	tgtcctctag	ttttgagaca	gctgcctcaa	cctctgcaga	acttcaagaa	1200
gccaataaaa	cattttggag	gataatcacc	ccaagtggc	ttcaaccata	aactttgtga	1260
ttccaaagtg	cccagttgtc	acaggtgcc	ttaaataatta	cattttccaa	cataaaaaaa	1320
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaagggc	ggccgc	1376

&lt;210&gt; 67

&lt;211&gt; 2434

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (10)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (12)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (27)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (73)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (75)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (103)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (130)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 67

ctgggggtan	tncaagaacc	ctctgtngga	cttagatgtc	aagctctttc	ctttgggcag	60
cgtgttttcc	ttntncgagt	agtggtgctg	gtaaactaaa	ttngccggtt	cgctttccat	120

tctctgacan	ttgagatgga	atgccttgac	cattggtgct	ctgacagaga	agtcatggag	180
tcattgccat	ttcctggttg	cccttttgga	atgtgatcct	gttagtagag	gttttctagc	240
ttctactaag	atatttcttt	ccctaaccat	catacacttg	gcatgtttca	ttcccatctc	300
ctttcccctc	accttaaagg	agactacccc	tttgccccat	attgtcaacc	taattttctc	360
tcgtactctc	tctagtgaat	gatgtgctac	caagtatatg	ccaggctgtg	agaggattat	420
actgagtagt	agaaagaagc	taatttgaaa	taaaaattat	ttgtataatt	aagaaagcag	480
attagatgca	catgggtcaac	aggaagttga	ctgtatgtct	gctagttaga	ttcaaaacat	540
cataaagatg	atagcatgtc	aatatattag	cctagccatt	atgttagcct	ttgttaggtg	600
ggcagctttt	ctgctttttc	ccttcctctg	tggtgacaac	ggaggaaata	tccaacagaa	660
atacgtctaa	caggggaaat	gggatcatag	tttatatgca	tctgatttga	aaggagtatt	720
gaggaaggtt	ttcatatatg	atctatcttt	ggattaaaaa	gaacatttat	gaaatcaagc	780
cttctaacac	tgtttataat	tgagaagcaa	cagtaactcc	gtggacagca	atcaagctta	840
aaattgtaaa	taaatatggg	gataattcag	ttgttgcaaa	aaaagggcag	aattcagtag	900
aataaagtc	ttttctctta	caggatttaa	atgaggacag	agaacctcag	gtgttcttat	960
gctagtgtt	gctgagtgc	tactaagaaa	gcaattccaa	atagatgtat	acatctagag	1020
agagtgggtat	tagagattca	gtgtatgtat	ttatttacat	gagaggaaac	tggaaataaa	1080
tcccataaat	tattggaata	taatcccata	aattatcacc	ttttatgact	ggaaaaatatt	1140
tgccaatgaa	gaaatgggtc	gtaggtattt	gtcttaagat	ttttggctgt	ttaataaaaa	1200
tgtaacttta	acggtttctt	atagttgcct	ttataaagtg	tattgtctaa	aatatttttg	1260
tatcatgtgc	ctttgaaatt	tgacagctga	tttgggtgtt	ggatttctgc	ccagccattt	1320
atcagtatta	tcatttttatt	cagtagctgg	cagggtgtatt	agacaaacga	gacttaggta	1380
aggaatggaa	cctttcctgt	ggtttgactg	cacatcacac	cagaagactc	cagtatccct	1440
cattccagaa	tgaggaaaaa	gtattctaca	aagaacctaa	tcacctctgt	gaaatctatg	1500
ggatggaaac	agtgtggcct	taggagtcaa	atagtctctg	catgggtggg	aggatcatga	1560
tggaaatagt	gaatttctac	ttctagaagt	tgtgaaatag	gtcctgcact	tttgcagaat	1620
gtccttcttt	aaacctggct	tattccacag	ctgtagctga	taacatgacc	tggggcttag	1680
ctgctctagc	cctgggttct	tggagacctc	acactgcctg	gcccctggcc	atccacctaa	1740
ggactgcctg	ctttctggtc	acatgtggac	cttgatacga	ctaagcgggt	acatatgtgg	1800
ttgtgcaaaa	gctttctggt	taatgcatag	tgttaccgat	ttacatcttg	gttttctagt	1860
gcactatgtc	taggaggcaa	tatcctttta	aacagtgcct	tggctaagat	agatacttgt	1920
gaatcaaaga	tagcacagaa	atgaactaag	tatatcccat	ttggaattat	attttgatac	1980
tatttaaaat	ggtttcacct	gttaaagggc	caacagaact	cttggtttta	cttttgtaat	2040
tactgtacag	aaaatttcaa	gagtgtttga	gtgcttgta	tcagggtgtt	tccttaataa	2100
gtagggatat	gatcattttac	aggaattata	tatgaaaaaa	gtttttgaaa	tgtatttttg	2160
tgatgtgcta	tgttgagggg	aaaccaaata	tttatgattt	taaaacattc	gtatgaaaac	2220
attgtacaat	gtaatatgct	caactttctc	aattttttgc	taatttttct	aagatacatt	2280
aaaaatgttt	tatatttttt	tttaagttaa	atggacccag	taagaaaaat	aaaaatacca	2340
gaacataaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	2400
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaa			2434

<210> 68  
 <211> 1086  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (10)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (77)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (1056)

<223> n equals a,t,g, or c

<400> 68

ttgaaacacn	cttttgagca	ttaggtccca	gctccaccgc	ggtggcgggc	gctctagaac	60
tagtgatcc	cccgggntgc	akkaattcgg	cacgagcaac	ctgtagtgag	ttggctgtca	120
ctcagcagct	ctggaaaatt	acctgtctgc	actgaatttc	ctcatcagta	aaatggaaat	180
gattatagta	ctgaccttgt	aggattatc	taaaaatcag	agaagttcat	gcagcttaga	240
acagtgccag	gcacatggca	aatgctatgg	catacttaag	catcttcctc	tgtggtgcct	300
cgtcatcacc	atgtgattgt	gccttgcttg	tccctgtttc	acttttcaga	gggagaaaag	360
tggccaactt	taagaatcaa	aattctgatg	ttacttcggg	aaatgcatag	agccagagag	420
acacaatttg	acttagtatg	atccacatca	tcccctcagg	ctgaatagtg	gtggcatgca	480
catctatacc	aaaatgtttt	accttttttg	tagaaggaaa	atatttgtat	cttctattcc	540
atatcttaga	tctttataag	agcacttaag	ttcaacctcc	taagaaactg	ccaattttgt	600
tgatcatgat	agtctgcaca	gattttcgtg	ctatttagtg	ktgggagtg	cttagggacc	660
atcaacaaca	ggscctcct	tttatccatg	agactactga	ggccttgagg	gttatttgtc	720
catccatggc	gggtgcacrg	ctagagggtta	tctggtagt	agcccaacta	agattagaac	780
ccaggaatth	tgatttaatc	tcaaatgacc	tcttttattt	ctgcatccgg	gaaggagaag	840
aggaaagtaa	acgggaaatt	catgttatct	atggaaaagt	tatcagtttg	atgtttatta	900
aatgatttgt	ttcaggagat	tgtttacaac	ttttatcttc	ctagacaata	attttctgta	960
agagtaaggt	catggtcatt	aaggtagtca	tattaatgta	ttcagtaasc	tgtgaagaaa	1020
aatatatatc	aatgttttcc	aataaaatc	agtgantacc	tgaaaaaaaaa	aaaaaaaaaa	1080
aaaaaa						1086

<210> 69

<211> 1262

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (568)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (639)

<223> n equals a,t,g, or c

<400> 69

cctyctatag	gkmaagctgg	tackcctgca	ggtaccggtc	cggaattccc	gggtcgaccc	60
acgcgtccgc	ttacaacttc	tctyttgttt	ctgtgtttct	tccacacccat	tatggcgat	120
ttactgatct	acaaccgggt	ttaagttagt	tgccctgctta	aaattttctc	atgaattctc	180
attacattht	caaacaaaat	ccaaagttht	taccatgacc	ttgtggatat	ctttctgacc	240
tcatttacta	ctgttatctt	ccttattcat	tccattccac	ccacagtggc	cttcttgcaa	300
gtcagagaat	gcaacaaaat	tgctcctgcc	ttggggcatt	tgcaattctg	ttcttttggc	360
ctggtcactt	cacctccaca	ttctccattt	tttatgactt	tcttccctata	tttgatctc	420
tgttcaaatg	tcacccctca	aaaaggcctt	caaaaactacc	ctatttataaa	tagcacctgc	480
cactgccatc	tccataccca	ttttcatttt	tccatagcac	ttatccctac	ttggcattat	540
gttatatatt	tgtttatttg	kttactgnct	gtctccccc	ctggactgta	agctccatga	600
ctgtaagtac	atagatgaat	tacaaaatga	atcaatggng	aawtatcctt	gmcatatatt	660
atttaattgat	tttgmtccca	ttttaagtaa	aaaaaaaaata	ccattttttt	cactcttcaa	720
agtgatatag	tttaattctc	taaactacat	ttttctcatt	tcccgattha	attaaatcag	780
tgatataaaa	aaaagagtga	tggggatatg	tgaaaagaaga	ctaaaataga	tgccaggaaa	840
tactaaaact	gctgaagttt	agtggtatat	ttttcttcta	cacacagtat	tatttgagtt	900
actaattgtg	tactgaatt	acagaataag	caaaaacta	ggtaaacaga	atcacgcttg	960
ggggctatat	tttgtgtaaa	atttgtgtta	tgcaaaaata	atattaaata	tttaattact	1020
acagttttgt	tatttctttc	ttatttttagg	aaatgatttg	cagctgagtg	aatcagggaag	1080
tgacagtgat	gactgaagaa	atatyttagct	ataaataaaa	atttatacag	catgtataat	1140

ttatatttgta ttaacaataa aaatttcctaa gactgagggga aatatgtctt aactttttgat	1200
gataaaagaa attaaatttg attcagaaat ttcaaaaaaa aaaaaaaaaa aagggcgggcc	1260
gc	1262

<210> 70  
 <211> 1642  
 <212> DNA  
 <213> Homo sapiens

<400> 70	
ggcgcgctgc cccgaretgc ctgggttgcg ctgccggcca cgtccccgcg cccggcctca	60
ggctccttcc tactgtccga gggccaccag gccgccgggg gctgtctgcg cccggatggc	120
tctgttacta gagtggagag tctaccttcg tctcacatgt gccacaaagg atggcatggc	180
ccgggagtgcc cccaccacgt ggctttcacc ccctgcaaag ccagacttcg cccagcgaca	240
cagtgtcaag cccacagctc tccaaggagg aagatgggtcc aggtctgggag catcycctta	300
gcagcagcct ctgatccctt ggccaagcag gagggaacca ttagcagcct gaggagctgg	360
ctggctggga gcctcgggga ccgcccagcc ttgctcccag ctcaccacaa agatgtggac	420
agctcttctg ctcatattga ttttytcctt gtccctatct gaaagccatg cggcatccaa	480
cgatccacgc aactttgtcc ctaacaaaat gtggaaggga ttagtcaaga ggaatgcac	540
tgtggaacaa gttgataata aaacgtctga ggatgtaacc atggcagcag cttctcctgt	600
cacattgacc aaagggatcc ggcagcccam ctcaactcta tggaagtcac aacagaggac	660
acaagcagga cagatgtgag tgaaccagca acttcaggag gtgcagctga tgggtgtgacc	720
tccattgctc ccacggctgt ggctccaggt acgactgcgg cctccattac gactgcggcc	780
tccagtatga ctgtggcctc cagtgtctcc acgactgcag cctccagtag aactgtggcc	840
tccattgctc ccacgactac agcctccagt atgactgcgg cctccagcac tcccatgaca	900
cttgcaactcc ccgcgccac gtccacttyc acagggcgga ccccgccac taccgccact	960
gggcatccat ctctcagcac agccctcgca caagtgccaa agagcagcgc gttgccaaaga	1020
acagcaaccc tggccacatt ggccacacgt gctcagactg tagcgaccac agcaaacaca	1080
agcagcccca tgagcactcg tccaagtcct tccaagcaca tgcccagtga caccgcggca	1140
agccctgtac cccctatgck tccccaaagca caaggtccca ttagccaggt gtcagtggac	1200
cagcctgtgg ttaacacaaac awataaatcc acamccatgc cctcaaacac aaccmcwgag	1260
cccctcacc aggccgtggt agacaaaact ctccttctgg tgggtgctgtt actcgggggtg	1320
acccttttca tcacagtctt ggttttgttt gccctgcagg cctatgagag ctacaagaag	1380
aaggactaca cccaggtgga ctacttaatc aacgggatgt atgcggactc agaaatgtga	1440
ggggggcggg ggcctggcgg gaggcctggc cccttcctcg tcccttcctt ttgcctttga	1500
gaccaaacca agtgcctcca aattcttttg gtgcaattga ggagatatgc cagatgctta	1560
aacacattta attgctgtca gattaattcc atgatcacta aagagttgct gcttttttca	1620
taaaaaaaaa aaaaaaaaaa gg	1642

<210> 71  
 <211> 921  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (4)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (9)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (11)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (15)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (20)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (901)

<223> n equals a,t,g, or c

<400> 71

gcnnggggna	naggnaagcn	ccccactatt	gggttcaaaa	gctggagctc	caccgcggtg	60
gcgggcgctc	tagaactagt	ggatcccccg	ggctgcagga	attcggcacg	aggtctgagc	120
agataagatt	aagggctggg	tctgtgctca	attaactcct	gtgggcacgg	gggctgggaa	180
gagcaaagtc	agcgggtgct	acagtcagca	ccatgctggg	cctgccgtgg	aagggagggtc	240
tgctctgggc	gctgctgctg	cttctcttag	gtccccagat	cctgctgata	tatgcctggc	300
atttccacga	gcaaagggac	tgtgatgaac	acaatgtcat	ggctcgttac	ctccctgcca	360
cagtggagtt	tgctgtccac	acattcaacc	aacagagcaa	ggactactat	gcctacagac	420
tggggcacat	cttgaattcc	tggaaggagc	aggtggagtc	caagactgta	ttctcaatgg	480
agctactgct	ggggagaact	aggtgtggga	aatttgaaga	cgacattgac	aactgccatt	540
tccaagaaaag	cacagagctg	aacaataact	tcacctgctt	cttcaccatc	agcaccaggc	600
cctggatgac	tcagttcagc	ctcctgaaca	agacctgctt	ggagggattc	caactgagtga	660
aaccactca	caggcttgct	catgtgctgc	tcccacattc	cgtggacatc	agcactactc	720
tyctgaggac	tcttcagtgg	ctgagcagct	ttggacttgt	ttgttatcct	atcttgcatg	780
tgtttgagat	tcagatcag	tgtttttagaa	aatccacaca	tcttgagcct	aatcatgtag	840
tgtagatcat	ttaacatcag	cattttaaga	aaaaaaaaaa	aaaaaaarct	cgagggggggg	900
nccggtaccc	agggcggaag	a				921

<210> 72

<211> 906

<212> DNA

<213> Homo sapiens

<220>

<221> SITE

<222> (34)

<223> n equals a,t,g, or c

<220>

<221> SITE

<222> (833)

<223> n equals a,t,g, or c

<400> 72

ggaaattctc	cctcactaat	tggaacaaaa	gctngagctc	caccgcggtg	gcgggcgctc	60
tagaactagt	ggatcccccg	ggctgcagga	attcggcacg	aggggaagaga	gaggggaggg	120
tgagcagagg	acaggccggg	agttttccgg	gaacggagga	agagcagtgg	aggctgccag	180
gatgaggctg	ctgtgtggcc	tgtggctgtg	gctctccttg	ctgaaagtcc	tgcaggccca	240
gacccaacc	cccctgccac	tcccggcccc	gatgcagagc	ttccaaggaa	accagttcca	300
gggggaatgg	ttcgtcctgg	gcctggcggg	caacagcttc	aggccggagc	acagggcgct	360
gctgaacgct	ttcaccgcaa	cttttgagct	aagtgatgat	ggccgctttg	aggtgtggaa	420

tgcgatgact	cgaggccagc	actgtgacac	atggtcttat	gtgctgatac	cggcagccca	480
gcctgggcag	ttcactgtgg	accacggtgt	gggcaggagc	tggttgctgc	ctcccgggac	540
gctggaccag	ttcatctgcc	tgggcagagc	tcarggcctc	tcggatgaca	acattgtctt	600
cccagatgtg	actggargtg	ccctggacct	carcagcctg	ccctgggtgg	cagccccage	660
ctgaccactc	agacagccgc	ggcccccaag	gcctgactct	tcttggtgga	gggcgaggct	720
ggtcacccca	ggccagcgtc	tggtgaagga	tgaagcagct	cctgtccggc	ccagccctgc	780
ctcacagctg	tgcgagctct	gccctcctca	gctctcaaac	ctgaataaat	gcnccaagcc	840
cagaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	ctcgaggggg	ggccccgtac	ccaattcgga	900
agattg						906

<210> 73  
 <211> 680  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (7)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (9)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (15)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (16)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (22)  
 <223> n equals a,t,g, or c

<400> 73	
cactcantng	aacannagct
cnagctccac	cgcggtggcg
gccgctctag	aactagtgga
60	
tcccccgggc	tgcaggaatt
cggcacgaga	tatttcgctg
gacctgtggg	ccatgatgct
accccaatgg	ctgctgctgc
120	
ctcttctctc	tcaccagggg
ctcactttct	ccaacaaaat
acaacctttt	ggagctcaag
240	
gagtcttgca	tccggaacca
ggactgcgag	actggctgct
gccaacgtgc	tccagacaat
300	
tgcgagtcgc	actgcgcgga
gaaggggtcc	gagggcagtc
tgtgtcaaac	gcaggtgttc
360	
tttgccagct	atagagcgtg
tccctgcctg	cgggaacctga
cttgatatata	ttcaaagaat
420	
gagaaaatggc	ttagcatcgc
ctatggccgt	tgtcagaaaa
ttggaaggca	gaagttggct
480	
aagaaaatgt	tcttctagt
ctccctcctt	cttgctgsct
cctcctycty	cacctgtctt
540	
cctccctacc	cagagctctg
tgktcacctt	gttccccaga
gcctccacca	tgagtggagg
600	
gaagtgggga	gtgattgaaa
taaagagctt	tttcaatgaa
aaaaaaaaaa	aaaaaaaaaa
660	
aaactcgagg	gggggcccgg
680	

<210> 74  
 <211> 1633  
 <212> DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 74

ggcaccgagca	tcagtaagta	ggagctaggg	aagagaaaagc	atgcaaaaagg	gcaagaatca	60
ggagaattga	tatagtagct	ggacagaact	ctagatgtgt	gtgtgtgaga	gaaagagagg	120
cagggagaag	gagggaggag	ttacccccac	gatgacctcc	aacttcacct	tctgcacct	180
catcctgggg	atagcacagg	ctcaggcctg	ccctgggttg	cctggcgatt	ggcctggcct	240
gggctcaggg	gtgggggagg	ggctgcacca	cattaggacc	tgccgtactc	caatcccatg	300
cagtcctcct	gctcctgctg	ctgcgtgcct	gggctctggg	catgccaggc	ttccttgtgt	360
cctgcgtctg	tggccgggttc	ctgccaacct	gtctagtcct	ttcaggcctg	aggccctgca	420
ttgctctttc	tggctccttc	ccctccttcc	cgctcccat	ctagcctttt	ttgggttccg	480
ggatctgctg	acagactttc	ttcttgctgc	ctgcctgctt	acatttcaga	agacccctct	540
ggaactgccc	atggctgtgg	tccacctgct	ggtagcaacg	ccctgttacc	aaatgctaga	600
taatctgcca	ctccctctg	cagccgcca	ctgggtgtga	gctgccacct	gtgtctgtt	660
tccacacctt	tgtacactgc	attctgcctg	tcactctgga	cagctgcca	gcagctgcag	720
ctgaggccgt	cctgccagca	cctaccgcag	ccccatgggt	ggctctgtcc	agttcctggc	780
gctcccagcc	tgcctcatct	gctgtactta	cagagccaca	tctgggtgaa	tacagagctg	840
gggtagcctg	gatgggcccc	tggcacactc	actcccagag	aggcgggcat	cctctctctg	900
gacctctagg	gaagggcagg	gtctggagcc	caataataaa	acacagtaat	gataatcata	960
gctaattgtt	actgagaact	taggatgtga	taggctcaat	gatttaccag	aattactcta	1020
tactgcataa	ctaccctatc	aggtaagtac	tattaatatc	ccctttccca	catgaagaaa	1080
ctaaggctaa	ccaatagaaa	tgaacctgtg	taatgtccca	tgggcataat	tcatggagcc	1140
aggattcaga	accagggtgct	ttacttcagg	gctgaagctc	ataaccacta	gaccaaagcg	1200
cctcctccgt	gacctctggg	gaaggccag	caacatcctc	tattgcgtcc	aatgacttct	1260
ccctgggtctg	agatccactg	actcacaggg	gtagggttaa	ggttaaggcc	agagtctcag	1320
ctaagctttg	gatactttct	tctggatttg	gagaaggctg	gaataactga	atttctgtct	1380
atcttcagga	gcgccctact	agagccacac	atttcccagc	tctccttggt	tgtttcccag	1440
acccttcttc	catcgtgtcc	tctcccttag	gctcctggaa	agttttcaga	gagaatcacc	1500
cagtggtaac	attgttaaac	aaaacaggaa	aatgggactt	gtgtgtatat	atgattaaat	1560
tattaattga	tggatctacc	ttcttagctc	gtgccgaatt	cgatatcaag	cttatcgata	1620
ccgtcgacct	cga					1633

&lt;210&gt; 75

&lt;211&gt; 1022

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 75

ttcccgggtc	gacccacgag	tccgcccacg	cgtccggctt	ggggccagca	ccctgtctca	60
aagatggcaa	aatgaggcta	gttctggatg	agctagctgg	tgtgggttcc	aaccatagga	120
acacactgat	gctcaaatec	taagggtcca	agctctaggc	cctggaggct	ggtagaacag	180
gatctatgcc	tggaaatcctg	gcagggatcc	ctgtcaagga	cttgtgttta	agcctgcttc	240
agggttctag	gctgcttctg	ctctgtgtct	gccagggctg	gctgagcggg	tggatgggtg	300
gacagaaggg	ctcaccaagg	attgtggaca	tagggtaggc	cctggtacca	cgggtttcag	360
gctgttatca	cttcccttgt	aggaacatag	ccagaagcag	atgagccagg	gtagagggtt	420
ggccccctct	ctcatcttcc	cttcagtctt	aaattgtctc	cagcgatggg	aagaggccag	480
ggactgtaac	ccttgtgctg	tgtattctct	gagcctctgc	tcactctcag	ggccaagcag	540
ctcccaagcc	ggggccctct	cttggccaaa	atctgaggag	cagtctaggt	tacaggcttt	600
ttggtaggta	ggttctggct	gcctgttaat	gcagttaggc	cccctgatta	ggtacagtga	660
gaaacaagct	agaacaaccc	tggcccagaa	gactgtgcac	tccagcaaga	tccagggatg	720
atagccttgc	agggccactg	ggagtgtgtg	cccaagcttc	tccctcttct	ctccccaggg	780
ggcactggga	ctggctcctg	ccctcatctt	tagcctgggc	cttccccaga	ggtattaaag	840
agaagtatga	ttcctctgtc	ttcagttctt	ctcaggggca	tcctgcccct	agtaccagct	900
tcccaagggg	ccccagtc	cgtgggtgaag	ctagcactc	atgcagctct	tagggaacca	960
aaaaccagca	ctgaaataaa	gctgaatgac	tgactgaaaa	aaaaaaaaaa	aaagggcggc	1020
cg						1022

<210> 76  
 <211> 1184  
 <212> DNA  
 <213> Homo sapiens

<400> 76  
 agcaaaaggt gaagagagaa aggaagcttt ttctctaaaa atgggtgcagc tatectctga 60  
 accaattttcc ttgggttttaa tgtacctgta tcttgggggtt tttttccact taatttatcc 120  
 tggagctctt tccataacaa cacttggaaa gcactctcat ctttttttca ctgctgaaca 180  
 gaattccact gtgtggatgg aacatactct atttcaccag tcccctgtag ccagtcactt 240  
 ggtttgtttc caatcttttg ctttttcaga gtaataacct tgtatgtcta tcattttgta 300  
 tgcatacagg tttatatgta ggaaaaattc ctagagtagg attgctggac caatggataa 360  
 aagtatatgt tggacagaca atgccaaatt gccttttcaga gactgtggcc ctgtgcaccc 420  
 catcaggcat gtgtgactac caaagctcct gtcagctgtt ttattttatc tcctttccag 480  
 tctcaggctc aatgcagAAC tttgaggtaa gctttttctaa aatgtaggct cctaaacgcc 540  
 acagccagct ctgccacatg aaggagagct caaatgagac agaaacagcc tctgggcagg 600  
 atttctatcc tgcacagata tattttccac attctgggaa accgtgaagc ttccagagcc 660  
 acaattcccc agaaacacat cccccgtgtg gtacagccaa gcccagaac aagctgtgct 720  
 tgccctggcac ctcaaagcca agcaccatgg atgccacttg ccatgggtgc ctgcaatttc 780  
 aaataatgag aaataagaaa tttcagcttc tcagtccctc tagccaacat ttcagggtca 840  
 tgacagcctc aggtggcaag cagctactct gccggacagg gcagaagatg gaacacccca 900  
 tccctgggga agatcacgag gtcaggagat caagaccatc ctggctaaca cagtgaacc 960  
 ccgtctctac tgaaaatata aaacattagc tgggtgtggt ggcaggcacc tgtagtccca 1020  
 gctactcagg aggtgagggc aggagaatgg cgtgaacctg ggaggtggag gttgcagtga 1080  
 gccaaagattg cgccactgca ctccagcctg gcaacagagt aagactccat atcaaaaaaa 1140  
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaa 1184

<210> 77  
 <211> 312  
 <212> DNA  
 <213> Homo sapiens

<400> 77  
 aattcccggg tcgaccacag cgtccgtgat gagtggattt gtactcttac ccaggctctg 60  
 agggccagcc caccacagcat ccccaccct gatgacgctg tccctacaac tggctgaact 120  
 ggtgcatttt gtgtgtgcct tccagagcca gtggactggt gtgtatccaa tgatgccacc 180  
 tctgaaacct acagaaccac tatgcttttg atgtgtaccc tgcagggtct gagggccagg 240  
 ctgtctggta gctctgctcc tgggtgacag agcaagactc tgtctcaaaa aaaaaaaaaa 300  
 agggcgggccg ct 312

<210> 78  
 <211> 1370  
 <212> DNA  
 <213> Homo sapiens

<400> 78  
 tggttaaaga gtacgacatt ttagtcaggg ctgagcacag cgcctggccg aggagtgcct 60  
 ttatatgtct ttggaagtat atggggactc ctttctgtc actgtggcga ttccacttat 120  
 gcattcacca tgccaatcca aggcagctga tggcttagga aagtcagaga ctgagatggt 180  
 aaaaaccttg gggctatcta ccaacatgtc tccattccac ctgttaggggt taaagggttt 240  
 tctaacctgg gccctgacct tagcacagat ctgcctatat ttttttgaag ttcagccact 300  
 tggactatta gccctaaact ttttctgtac tggcactgca gggctgaagg agctttgcat 360  
 gcatccacca agtctggctt tcacacctga atttcacacc tcgctttcac ccttagctat 420  
 tccatctttc tgtggaacat cagtgtcact tagcaatagc catacaatcc cattatcctt 480  
 atacctacct ttcccttcaa agtctcggat gcctgatata ttgcacctgc tagtgcattc 540  
 actcccatta gtacactccc aagtccttcc agtgaaagat gtaacaattg aatggccact 600  
 gtgccaaagg tgcctgggct ctacctgcca ccagtgatgg ggtcctcaca gccaaacccag 660





<220>  
 <221> SITE  
 <222> (1)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (4)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (5)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (9)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (11)  
 <223> n equals a,t,g, or c

<400> 80  
 ntttnnaacng naaatctccc cttactattg ggaacaaaag ctggagctcc accgcggtgg 60  
 cggccgctct agaactagt gatcccccg gctgcaggaa ttcggcacga gattccccag 120  
 aattgggttca tcttgggaaa ggaaggccta ggaagctgat gacctacttg ttttgttcat 180  
 ccatttctct actgcttctt aaagtgcact caagcgggtca ccaggacatc agaaaagcca 240  
 aatccaaggt cccaaggctc ttgatcatcc agtgcccaca acagaggggag taagctagac 300  
 ggagaagctg aagtgaagga ggaaagagag aaggggggtgc ctgttccacc cctttcctag 360  
 cagctgcctc tgtgggctcc ataatcccc tgcctctcc cccaccccac cctccggccg 420  
 mcccactctc ctccacccac cccattcatc agtaggaggg cactgtatca gcacagcctt 480  
 tcaaggcctg acagttcctt tctgggtgcc cccacatgt aaccacttaa cccccagcac 540  
 aaagggccag cctgaatggg cttggagccg gtggctcgtg cgraggggga ggcctcaata 600  
 ggattttgggg agctgggtgt ataaaccgct ctgctccagc agccctcaat ggaggagctg 660  
 gagtgggggc atgtgccgag gggaggagct cggcagctgc atcccgscct tgaacaatga 720  
 tttattcatt ctctggcac gcacagagaa gacaggctgg ggaaacatgg cccacaaaat 780  
 cgccctcagc ctccagctgat tcccacggca atcattgcag ttcagagtgc cccctccac 840  
 ttgtaatgtt ttgcctcaag cttgcccttc ctcttctcgg taaggaagta gactacatta 900  
 cgggaggcct taagtagcct taaattcagg caagcatcgt tctttgagtt ggttgacccc 960  
 aaactcaatc tcaactctct aaagctttct ggccagggtt tcctcacaca cccatagaaa 1020  
 cagatgcaga caaaaaaaaa aaaaaaaaaa actcgagggg gggcccggaa cccaattcgg 1080  
 aagatttg 1088

<210> 81  
 <211> 1862  
 <212> DNA  
 <213> Homo sapiens

<400> 81  
 ggcacgagct ctgcaaagct ccctctgctc cagcacgtcc tgtatttttt caactttgga 60  
 tgtttgcaca agctattatc ttttgcaaga atgtcctgcc ccctcctatc catctgtgga 120  
 agtcaacctt tctttagggc ccagaggaaa acctatcacc ttctacacct tcccagatgc 180  
 ccaccatttg ggttaaactc tgcctctctc aggtttgcca tgggtctttc cctctcttaa 240  
 agcactgggt ccagcctatg ccattgtgtg tcaccttggc tctgtctcc tactggctgt 300  
 agatcttttg gargggagga actgggggaat tttcatcctg gggctatggg atagtgtta 360

cccctaagta	taaaataagc	aagaatcctt	gaattgaaaa	gctccagtga	cttgatagct	420
tccaacagcc	cctaaagatg	gggtagtcca	ctgctcccta	tataggaagg	aagaggttca	480
cagtctcatg	gatttagcag	ccttcgtagc	ctcaggatct	ttcccaatgt	ccccagttct	540
cctgctgggc	cagctgtggt	ctctgagaat	gctcagttcc	ctgggctagg	ctggcttatg	600
gggcccctcc	tcctggacct	gggtccatgt	gaccaagtgg	ccagccagcg	agagagccct	660
gggtgatgag	catctggaag	tcagcttgct	tgtctgccct	aatcacagcc	ctgggctcta	720
ccctcaagta	ccctactcac	atgggtccaag	agggaaaccac	tcttctgtca	ttccctcccc	780
ttctgctcct	catcctaccc	acccagctgt	tctgggcaag	atcctgtggg	catgatggaa	840
tcctcccttt	tctctacccat	ccatcccttt	gatcacctaa	tcctrcarct	gccctccaac	900
caaggtgttc	tttctatacc	aaaaaatctg	atcaggccat	cactcctctg	tttaaaacca	960
ttcaacggct	acccactgcc	ctcaggataa	aaactaaaat	tcttaacagg	atctacaaat	1020
ttctcacctg	gcccttttga	gataagagga	tttataatct	tttcaaagcc	aacaagcatc	1080
ctgagacaat	cttaattttt	gcagctgtcc	aagctattta	agattgatga	ggtctgcamt	1140
aagttgttaa	atgggtgatta	atttatctgg	cctgaatggc	akgtgaattc	caaggggatc	1200
atctttctct	accttccttg	ctgcacattg	ggacacagca	aggccaggac	acccttytta	1260
ggtgactgsg	taaccagggt	cattaaggca	gtcagcgag	agaagaagca	gaaaatgggg	1320
aggagttgtg	taaatgaaat	tgccgataat	accaaaggaa	ccccaagaga	ggttcagtct	1380
cccaacaatt	gcagctttgc	tgcaacatgg	caggggtggga	caggggtagg	tttggtatctg	1440
ggctgctttg	ccmtggaagt	ggggttacag	ggcagggaga	ggcttaccat	cactgagcat	1500
ctttaattga	ctttgtcatt	ttastgagaa	gttacaatct	gagctattct	tcattgagtt	1560
atattttgaa	gcaggttcaa	tacataatac	agtccctcct	ctagaatttc	tgctttatgt	1620
ccacgtgatg	cagcttggtc	atcttctgaat	tcctttttaat	gacaaaaaca	tttggaagga	1680
atgtaaagac	tcaaaaaggg	gatggcatgt	aaactatagt	tggacaata	tgtggccatt	1740
taaaagaagt	gttgtccagg	aggtggagg	tgcagtgagc	tgagattgca	ccactgcact	1800
ctagcctggg	ccatagaatg	agactccttc	tcaaaaaaaa	aaaagaaaaa	aaaaaactcg	1860
ag						1862

&lt;210&gt; 82

&lt;211&gt; 1618

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 82

ccacgcgtcc	gcaaggagcc	agaggccatg	cagtggctca	gggtccgtga	gtcgcctggg	60
gaggccacag	gacacagggt	caccatgggg	acagccgccc	tggttcccg	ctgggcagcg	120
ctcctgetct	ttctcctgat	gtgtgagatc	cctatgggtg	agctcacctt	tgacagagct	180
gtggccagcg	actgccaaag	gtgctgtgac	tctgaggacc	ccctggatcc	tgcccatgta	240
tcctcagcct	cttctccggg	ccgccccccac	gccctgcctg	agatcagacc	ctacattaat	300
atcaccatcc	tgaagggtga	caaaggggag	ccaggcccaa	tgggcctgcc	agggtacatg	360
ggcagggagg	gtccccaaag	ggagcctggc	cctcagggca	gcaaggggtga	caagggggag	420
atgggcagcc	ccggcgcccc	gtgccagaag	cgcttcttcg	ccttctcagt	gggccgcaag	480
acggccctgc	acagcggcga	ggacttccag	acgctgctct	tcgaaagggt	ctttgtgaac	540
cttgatgggt	gcttttgacat	ggcgaccggc	cagtttgctg	ctcccctgcg	tggcatctac	600
ttcttcagcc	tcaatgtgca	cagctggaat	tacaaggaga	cgtacgtgca	cattatgcat	660
aaccagaaag	aggctgtcat	cctgtacggg	cagcccagcg	agcgcagcat	catgcagagc	720
cagagtgtga	tgtctggacct	ggcctacggg	gaccgcgtct	gggtgcggct	cttcaagcgc	780
cagcgcgaga	acgccatcta	cagcaacgac	ttcgacacct	acatcacctt	cagcggccac	840
ctcatcaagg	ccgaggacga	ctgagggcct	ctggggccacc	ctcccggctg	gagagctcag	900
ctgatacggc	atcctgcgag	aagacctgcc	ctcctcactg	ggatccctct	cctgcctcct	960
cccagggtct	tgccaggggc	ttgctcagtc	ccttccacca	aagtcatctg	aacttccggt	1020
tcccaggggc	tccagctgcc	ctcagacact	gatgtctgtc	cccagggtgt	ctctgcccct	1080
catgccccct	tcaccggccc	agtgcctcga	ctctccaggc	tttatcaagg	tgctaaggcc	1140
cgggtgggca	gtcctctgct	tcagagccct	cctccggcct	ggtgctgcct	ttacaaacac	1200
ctgcaggaga	agggccacgg	aagccccagg	ctttagagcc	ctcagcaggt	ctggggagct	1260
agagcaaagg	agggacctca	ggccttccgt	ttcttcttcc	aggggtgggt	ggcctgggtg	1320
tcccctagcc	ttccaaaccc	aggtggcctg	cccttctccc	cagagggagg	cggcctccgc	1380
ccattgggtg	tcatgcagac	tctggggctg	aggtgccccg	gggggtgatc	tctggtgctc	1440
acagtcgagg	gagccgtggc	tccatggcca	gatgacggaa	acagggtctg	accaagtggc	1500

aggaagacct gtgctataaa ccaccctgcc tgatcctgcc cctgcctgac cccgccacgc 1560  
 cctgccgtcc agcatgatta aagaatgctg tctcctcttg gaaaaaaaaa aaaaaaaa 1618

<210> 83  
 <211> 2034  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (14)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (382)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (1999)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (2027)  
 <223> n equals a,t,g, or c

<400> 83  
 actcactaag ggancaaaagc tggagctcca ccgcggtggc ggccgctcta gaactagtgg 60  
 atcccccggg ctgcaggaat tcggcacgag atctaccggg agakggaytc tgaaagggcc 120  
 cgggccagcg tccctgagac gccaacggca gtcactgccc cccattccag ctcttgggat 180  
 acgtactatc agccccgtgc cctggagaaa catgctgaca gcacccctggc actggcttca 240  
 gtattctggg ccactctctta ttactcctct cccttcgcct tcttctactt gtacaggaaa 300  
 ggttacttga gtttgtccaa agtggtgccg ttttctcact atgctggggac attgctgcta 360  
 cttctggcag gtgtggcctg cntccgaggc attggccgct ggaccaaccc ccagtaccgg 420  
 cagttcatca ccactcttga agcaacacat cgggaaccagt cttcagaaaa caagaggcag 480  
 cttgccaaat acaactttga cttccggagc tggccagtcg acttccactg ggaagaaccc 540  
 agcagccggg agcagtcctc agggggccct tcccgcggg gtgtggccct gcttcgcccc 600  
 gagccccctgc accggggggc agcagacacc ctccctcaacc ggggtaagaa gctgccttgt 660  
 cagatcacca gctacctggg ggcgcacacc ctaggggcgc ggatgctgta tccaggctct 720  
 gtgtaccctgc tgcagaaggc cctcatgcct gtgctgctgc agggccaggc ccgactgggtg 780  
 gaagagtgtg atgggcgcgg ggcaaagctg ctggccctgt atggcaatga gattgacacc 840  
 atgtttgtgg accggcgggg gacagctgag ccccaggggc agaagctggg gatctgctgt 900  
 gaggggaatg ctgggtttta tgaggtgggc tgcgtctcca cgccccctga agctggatat 960  
 tcagtccctg gctggaatca tccaggcttt gctggaagca cgggggtggc attcccgcag 1020  
 aatgaggcta atgccatgga tgtgtgggtc cagtttgcca tccaccgcct rggcttccag 1080  
 ccccaggaca tcatcatcta cgctgggtcc atcggcggtc tccactgccac gtgggcagcc 1140  
 atgtcctacc cagatgttag tgccatgatc ctggatgcct cctttgatga cctgggtgcc 1200  
 ttggccttga aggtcatgcc agacagctgg aggggcctgg tgaccaggac cgtgaggcag 1260  
 catctcaatc taaacaacgc ggagcagctg tgcagatacc agggtcctgt actgctgatc 1320  
 cggagaacca aggatgagat catcaccacc acggttccct aggcacatcat gtccaaccga 1380  
 ggcaatgacc tctctgtgaa gctcctgcag catcggtatc cccgggtgat ggcagaggag 1440  
 ggtcttcgag tgggtaggca gtgggtggag gcctcctcac agctggagga agcctcaatt 1500  
 tatagccgat gggaggtgga agaggactgg tgtctgtctg tcttcgctc ctaccaggca 1560  
 gaacacgggc ccgacttccc ctggagcgtg ggggaggaca tgagtgcaga tggacggcgg 1620  
 cagctggctt tgtttctggc tcggaagcat ctgcacaact ttgaggccac tccactgcacc 1680  
 ccactcccag ccagaaactt ccagatgccc tggcacctct agggaccaac tgggactcat 1740

tatggaagaa	tggggtgaga	ggagacatga	ggaaagaccc	tcttatttgt	gattctctgt	1800
gttcatgttg	ctgtttatag	tttgtggaaa	gtgggggacc	atcccccttc	tcaccactgt	1860
tcctcttgca	cgtttcccct	cattcatgtg	gctgtactta	accttctcca	acatacatcc	1920
tgcattacat	gaatggatta	ttcctaataa	ttaataaaaa	ggtatttttt	ctacaaaaaa	1980
aaaaaaaaaa	aaactccana	attacttcct	aaaaccggcc	cgcggncccc	atcc	2034

&lt;210&gt; 84

&lt;211&gt; 2240

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 84

tcgaccacgc	gtccgcttcc	tcattgctgcc	cgctcgcccta	ccgttcaggc	tgctgagcct	60
tttcccttct	ggatccgctc	ccacggcagc	gcgccatggc	ctccggggagc	cgctcctgga	120
gaggaggtgc	gctgcgcct	cctccttcca	gcactcatcg	agtctgggac	gcgagcttcc	180
ttatgacccc	gtggacacgg	agggcttttg	agaaggtggt	gacatgcagg	agcgttttct	240
gttcccggag	tacatcctgg	atccggagcc	gcaacccacc	cgcgaaaagc	agctgcagga	300
gtcccagcaa	cagcaggagg	aggaggagcg	acagaggcag	cagcggcggg	aggagcggcg	360
acagcaaaac	ctacggggcca	ggccccggga	gcacccgggc	gtggggcacc	cggaacccggc	420
attgccgccc	agcggcgctga	actgctcggg	ctgcggggca	gagctgcact	gccaggacgc	480
cggagtgcgc	ggctacctgc	cccgagagaa	gttccctcgc	acggcgggagg	cagacggcgg	540
gctggcacgg	accgtgtgcc	agcgtgtctg	gctgctgtcg	caccaccggc	gcgctctacg	600
cctgcagggtg	agccgcgagc	agtacctgga	gctgggtgagc	gccgcgttgc	ggcggcccg	660
ccccctcctg	gtgctctaca	tgggtggacct	gctgggacctg	cccgcgcccc	tgctgcccga	720
cttgcccgcg	ctgtgtgggc	ccaagcagct	gatcgtgctg	ggaaacaaag	tggacctcct	780
gccccaggat	gctcctggct	accggcagag	gctgcgggag	cgactgtggg	aggactgtgc	840
ccgcgcgggg	ctcctgctgg	ccccgggcca	ccaagggcca	cagcgccccg	tcaaggacga	900
gccacaggac	ggggagaatc	cgaatccgcc	gaactgggtcc	cgcacagtgg	tcagggacgt	960
gcggctgctc	agcgccaaga	ccggctatgg	agtgggaagag	ttgatctctg	cccttcagcg	1020
ctcctggcgc	taccgtgggg	acgtctactt	agtgggcgcc	accaacgccc	gcaaattccac	1080
tctctttaac	acgtcctctg	agtccgatta	ctgcactgcc	aagggctccg	acgccatcga	1140
cagagccacc	atctccccct	ggccaggtag	tacattaaac	cttctgaagt	ttcctatttg	1200
caacccaact	ccttacagaa	tgtttaaaag	gcatacaaaga	cttaaaaaag	attcaactca	1260
agctgaagaa	gatcttagtg	agcaagaaca	aaatcagctt	aatgtcctca	aaaagcatgg	1320
ttatgtcgta	ggaagagttg	gaaggacatt	cttgtattca	gaagaacaga	aggataacat	1380
tccttttgag	tttgatgctg	attcacttgc	ctttgacatg	gaaaatgacc	ctgttatggg	1440
tacacacaaa	tcaccaaacc	aagtagaatt	gactgcacaa	gatgtgaaag	atgccactg	1500
gttttatgac	acccctggaa	ttacaaaaga	aaattgtatt	ttaaattctc	taacagaaaa	1560
agaagtaaat	attgttttgc	caacacagtc	cattgttcca	agaacttttg	tgcttaaac	1620
aggaatgggt	ctgttttttg	gtgctatagg	ccgcatagat	ttcctgcagg	gaaatcagtc	1680
agcttggttt	acagtcgtgg	cttccaacat	cctccctgtg	catatcacct	ccttggacag	1740
ggcagacgct	ctgtatcaga	agcatgcagg	tcatacgtaa	ctccagattc	caatgggtgg	1800
aaaagaacga	atggcaggat	ttcctcctct	tggtgctgaa	gacattatgt	taaaagaagg	1860
actgggggca	tctgaagcag	tggccgacat	caagttttcc	tctgcagggt	gggtttcagt	1920
aacacctaat	tttaaggaca	gactgcattc	ccgaggctat	acacctgaag	gaacagtttt	1980
gaccgtccgg	ccccctctct	tgccatatat	tgtaaacatc	aaaggacagc	gcatacaagaa	2040
aagtgtggcc	tataaaacca	agaagcctcc	ttcccttatg	tacaacgtga	ggaagaagaa	2100
aggaaagata	aatgtatgag	accgaccttg	ttcactccag	atattaactg	tattgaacac	2160
aacaaaatac	attgaatttg	tattaaacat	ataacgcata	aataaagctc	ccattctttac	2220
ccttaaaaaa	aaaaaaaaaa					2240

&lt;210&gt; 85

&lt;211&gt; 1488

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 85

cgccaagttt	cgggagggag	agggtagaaa	ctggaggggg	tggacctgtc	actcacggga	60
ctgaggggtcc	ttttctcccc	ctcccaggag	gaacgagaat	gaatatgact	caagcccggg	120
ttctgggtggc	tgcagtgggtg	gggttgggtg	ctgtcctgct	ctacgcctcc	atccacaaga	180
ttgaggaggg	ccatctggct	gtgtactaca	ggggaggagc	tttactaact	agccccagt	240
gaccaggcta	tcatatcatg	ttgcctttca	ttactacgtt	cagatctgtg	cagacaacac	300
tacaaactga	tgaagttaaa	aatgtgcctt	gtggaacaag	tggtaggggtc	atgatctata	360
ttgaccgaat	agaagtgggtt	aatatgtttg	ctccttatgc	agtgtttgat	atcgtgagga	420
actatactgc	agattatgac	aagaccttaa	tcttcaataa	aatccaccat	gagctgaacc	480
agttctgcag	tgcccacaca	cttcaggaag	tttacattga	attgtttgat	caaatagatg	540
aaaacctgaa	gcaagctctg	cagaaagact	taaacctcat	ggccccagg	ctcactatac	600
aggctgtgcg	tgttacaaaa	cccaaaatcc	cagaagccat	aagaagaaat	tttgagttaa	660
tggaggctga	gaagacaaaa	ctccttatag	ctgcacagaa	acaaaagggt	gtggaaaaag	720
aagctgagac	agagaggaaa	aaggcagtta	tagaagcaga	gaagattgca	caagtggcaa	780
aaattcgggt	tcagcagaaa	gtgatggaaa	aagaaactga	aaagcgcatt	tctgaaatcg	840
aagatgctgc	attcctggcc	cgagagaaa	cgaaagcaga	tgctgaatat	tatgctgcac	900
acaaatatgc	cacctcaaac	aagcacaagt	tgaccccgga	atatctggag	ctcaaaaagt	960
accaggccat	tgcttctaac	agtaagatct	atcttggcag	caacatccct	aacatgttcg	1020
tggactcctc	atgtgctttg	aaatattcag	atattaggac	tggagagaaa	agctcactcc	1080
cctctaaggc	ggctcttgaa	ccctctggag	agaacgtcat	ccaaaacaaa	gagagcacag	1140
gttgatgcaa	gaggtggaaa	tgttctccat	atcaagatgt	ggcccaagg	gttaagtggt	1200
aacaatcatt	atacggactc	ttcagattta	cagagaactt	acacttcate	tgttccacct	1260
ctcctgcgat	agtcctgggt	gctccactga	ttggaggata	gagccagctg	tctgacacac	1320
aatgggtctt	ttcagccaca	gtcttatcaa	gtatcctata	tgtattcctt	tctaaactgc	1380
tactcatgaa	tgaggaaaag	ctgatgctaa	gatactgcct	gcactggaat	gttaaacact	1440
aaatatataa	caagctgtgt	tttcctaagc	tgaaaaaaaaa	aaaaaaaaa		1488

&lt;210&gt; 86

&lt;211&gt; 3174

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 86

gcgagcgtg	grcscaaaca	ctaaggcctg	agcggtgaca	atcgaggcga	gatgatgggtc	60
aacagggaat	gcctcgtggg	agaaaaaaga	caattttatt	ctcagcgtg	atcttgagat	120
gatgggcttg	ggaaacgggc	gtcgcagcat	gaagtcgccg	cccctcgtgc	tggccgccct	180
ggtggcctgc	atcatcgtct	tgggcttcaa	ctactggatt	gagagctccc	ggagcgtgga	240
cctccagaca	cggatcatgg	agctggaagg	cagggtccgc	aggcgggctg	cagagagagg	300
cgccgtggag	ctgaagaaga	acgagttcca	gggagagctg	gagaagcagc	gggagcagct	360
tgacaaaatc	cagtcagcc	acaacttcca	gctggagagc	gtcaacaagc	tgtaccagga	420
cgaaaaggcg	gttttgggtg	ataacatcac	cacaggtgag	aggctcatcc	gagtgctgca	480
agaccagtta	aagaccctgc	agaggaatta	cggcaggctg	cagcaggatg	tctccagtt	540
tcagaagaac	cagaccaacc	tggagaggaa	gttctcctac	gacctgagcc	agtgcacaa	600
tcagatgaag	gaggtgaagg	aacagtgtga	ggagcgaata	gaagagggtc	ccaaaaagg	660
gaatgaagct	gtagcttcca	gagacctgag	tgaacaaca	gaccagagac	agcagctcca	720
agccctcagt	gagcctcagc	ccaggctgca	ggcagcaggc	ctgccacaca	cagaggtgcc	780
acaaggggaag	ggaaacgtgc	ttggtaacag	caagtcaccg	acaccagccc	ccagttccga	840
agtgggtttt	gattcaaaga	gacaagttga	gaaagaggaa	accaatgaga	tccaggtggt	900
gaatgaggag	cctcagaggg	acaggctgcc	gcaggagcca	ggccggggag	aggtgggtgga	960
agacagacct	gtaggtggaa	gaggcttcgg	gggagccgga	gaactggggc	agacccccaca	1020
ggtgcaggct	gccctgtyag	tgagccagga	aaatccagag	atggagggcc	ctgagcgaga	1080
ccagcttgct	atccccgacg	gacaggagga	ggagcaggaa	gctgccgggg	aaggagagaa	1140
ccagcagaaa	ctgagaggag	aagatgacta	caacatggat	gaaaatgaag	cagaatctga	1200
gacagacaag	caagcagccc	tggcagggaa	tgacagaaac	atagatgttt	ttaattgttg	1260
agatcagaaa	agagacacca	taaaatttact	tgatcagcgt	gaaaagcgga	atcatacact	1320
ctgaattgaa	ctggaatcac	atattttcaca	acaggggccga	agagatgact	ataaaatggt	1380
catgagggac	tgaatactga	aaactgtgaa	atgtactaaa	taaaatgtac	atctgaagat	1440
gattattgtg	aaattttagt	atgcactttg	tgtaggaaaa	aatggaatgg	tcttttaaac	1500
agcttttggg	gggtactttg	gaagtgtcta	ataagggtgtc	acaatttttg	gtagtaggta	1560

tttcgtgaga	agttcaacac	caaaactgga	acatagttct	ccttcaagt	ttggcgacag	1620
cggggccttc	tgattctgga	atataacttt	gtgtaaatta	acagccacct	atagaagagt	1680
ccatctgctg	tgaaggagag	acagagaact	ctgggttccg	tcgtcctgtc	cacgtgctgt	1740
accaagtgt	ggtgccagcc	tgttacctgt	tctactgaa	aagtctggct	aatgctcttg	1800
tgtagtcact	tctgattctg	acaatcaatc	aatcaatggc	ctagagcact	gactgttaac	1860
acaaacgtca	ctagcaaagt	agcaacagct	ttaagtctaa	atacaaagct	gttctgtgtg	1920
agaatttttt	aaaaggctac	ttgtataata	acccttgtca	tttttaagt	acaaaacgt	1980
attaagtggc	ttagaatttg	aacattttgt	gtctttat	actttgtctc	gtgtgtgggc	2040
aaagcaacat	cttccctaaa	tatatattac	caagaaaagc	agaagcaga	ttagggtttt	2100
gacaaaacaa	acaggccaaa	agggggctga	cctggagcag	agcatgggtg	gaggcaaggc	2160
atgagagggc	aagtttgttg	tggacagatc	tgtgcctact	ttattactgg	agtaaaagaa	2220
aacaaagtcc	attgatgtcg	aaggatata	acagtgttag	aaattaggac	tgtttagaaa	2280
aacaggaata	caatgggttg	ttttatcata	gtgtacacat	ttagcttgtg	gtaaatgact	2340
cacaaaactg	atttttaaat	caagttaatg	tgaattttga	aaattactac	ttaatcctaa	2400
ttcacataaa	caatggcatt	aaggtttgac	ttgagttggg	tcttagtatt	atttatggta	2460
aataggctct	taccacttgc	aaataactgg	ccacatcatt	aatgactgac	ttcccagtaa	2520
ggctctctaa	ggggtaagta	ggaggatcca	caggatttga	gatgctaagg	ccccagagat	2580
cgtttgatcc	aaccctctta	ttttcagagg	ggaaaatggg	gcctagaagt	tacagagcat	2640
ctagctgggt	cgctggcacc	cctggcctca	cacagactcc	cgagttagctg	ggactacagg	2700
cacacagtca	ctgaagcagg	ccctgtttgc	aattcacggt	gccacctcca	acttaaacat	2760
tcttcatatg	tgatgtcctt	agtcactaag	gttaaaactt	cccacccaga	aaaggcaact	2820
tagataaaat	cttagagtac	tttcatactc	ttctaagtcc	tcttccagcc	tcactttgag	2880
tcctccttgg	ggttgatagg	aattttctct	tgttttctca	ataaagtctc	tattcatctc	2940
atgtttaatt	tgtacgcata	gaattgctga	gaaataaaat	gttctgttca	acttaaaaaa	3000
cttgtcacag	ctctgttttt	ttgtttttcc	ccttcatect	ctattagatc	cttaagaagt	3060
ccacctccta	actgtggact	aaggatggca	gtttgcctta	aaccttctgt	ggaatgagat	3120
gggctgtcag	gacgtcttca	accacagta	ccccgaaagc	tttgtgtca	ccac	3174

&lt;210&gt; 87

&lt;211&gt; 2780

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (2760)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 87

ggcgccgcgg	gcggggggtc	ccggggccac	gtggacgccg	acggctctgg	tctgcggcgt	60
gtggacactg	ccagcggctg	cagccgactt	ggaatgacct	gggagacaaa	tacaacagca	120
tggaagakgc	caaagtctat	gtggctaaag	tggactgcac	ggcccactcc	gacgtgtgct	180
ccgcccaggg	ggtgcgagga	tacccacact	taaagctttt	caagccaggc	caagaagctg	240
tgaagtacca	gggtcctcgg	gacttccaga	cactggaaaa	ctggatgctg	cagacactga	300
acgaggagcc	agtgcaccca	gagccggaag	tggaaaccgc	cagtgcctcc	gagctcaagc	360
aagggctgta	tgagctctca	gcaagcaact	ttgagctgca	cgttgcacaa	ggcgaccact	420
ttatcaagtt	cttcgctccg	tggtgtgggt	actgcacaa	cctggctcca	acctgggagc	480
agctggctct	gggccttgaa	cattccgaaa	ctgtcaagat	tggcaagggt	gattgtacac	540
agcactatga	actctgctcc	ggaaaccagg	tctgtggcta	tcccactctt	ctctggttcc	600
gagatgggaa	aaaggtggat	cagtacaagg	gaaagcggga	tttggagtca	ctgagggagt	660
acgtggagtc	gcagctgcag	cgcacagaga	ctggagcgac	ggagaccgtc	acgccctcag	720
aggccccggt	gctggcagct	gagcccagg	ctgacaaggg	cactgtgttg	gcaactcactg	780
aaaataactt	cgatgcaccc	attgcagaag	gaataacctt	catcaagttt	tatgtccat	840
ggtgtgggtca	ttgtaaagct	ctggctccta	cttgggagga	actctctaaa	aaggaattcc	900
ctgggtctggc	gggggtcaag	atcgccgaag	tagactgcac	tgctgaacgg	aatatctgca	960
gcaagtattc	ggtacgaggc	tacccacagt	tattgttttt	ccgaggaggg	aagaaagtca	1020
gtgagcacag	tggaggcaga	gaccttgact	cgttacaccg	ctttgtcctg	agccaagcga	1080
aagacgaact	ttaggaacac	agttggaggt	cacctctcct	gcccagctcc	cgcacctctg	1140

gtttaggagt	tcagtccac	agaggccact	gggttcccag	tggtggctgt	tcagaaagca	1200
gaacatacta	agcgtgaggt	atcttctttg	tgtgtgtgtt	ttccaagcca	acacactcta	1260
cagattcttt	attaagttaa	gtttctctaa	gtaaatgtgt	aactcatggt	cactgtgtaa	1320
acattttcag	tggcgatata	tcccctttga	ccttctcttg	atgaaattta	catggtttcc	1380
tttgagacta	aaatagcgtt	gagggaaatg	aaattgctgg	actattttgt	gctcctgagt	1440
tgagtgattt	tggtgaaaga	aagcacatcc	aaagcatagt	ttacctgccc	acgagttctg	1500
gaaaggtggc	cttgtggcag	tattgacgtt	cctctgatct	taaggtcaca	gttgactcaa	1560
tactgtgttg	gtccgtagca	tggagcagat	tgaaatgcaa	aaaccacac	ctctggaaga	1620
taccttcacg	gccgtgctg	gagcttctgt	tgctgtgaat	acttctctca	gtgtgagagg	1680
ttagccgtga	tgaaagcagc	gttacttctg	accgtgcctg	agtaagagaa	tgctgatgcc	1740
ataactttat	gtgtcgatac	ttgtcaaate	agttactgtt	caggggatcc	ttctgtttct	1800
cacgggggtg	aacatgtctt	tagttcctca	tgtaaacacg	aagccagagc	ccacatgaac	1860
tggtggatgt	cttctttaga	aagggtaggc	atggaaaatt	ccacgaggct	cattctcagt	1920
atctcattaa	ctcattgaaa	gattccagtt	gtatttgtca	cctggggtga	caagaccaga	1980
caggctttcc	caggcctggg	tatccaggga	ggctctgcag	ccctgctgaa	gggccctaac	2040
tagagtctta	gagtttctga	ttctgtttct	cagtagtctt	tttagaggct	tgctatactt	2100
ggtctgcttc	aaggaggtcg	accttctaatt	gtatgaagaa	tgggatgcat	ttgatctcaa	2160
gaccaaagac	agatgtcagt	gggctgctct	ggccctgggtg	tgcacggctg	tggcagctgt	2220
tgatgccagt	gtcctctaac	tcagtctgtc	cttgtgatta	aacacctcta	tctcccttgg	2280
gaataagcac	atacaggctt	aagctctaag	atagggtgtt	gtccttttac	catcgagcta	2340
cttcccataa	taaccacttt	gcaccaaca	ctcttcaccc	acctcccata	cgcaagggga	2400
tgtggatact	tgcccaaaag	taactggtgg	taggaatctt	agaaacaaga	ccacttatac	2460
tgtctgtctg	aggcagaaga	taacagcagc	atctcgacca	gcctctgcct	taaaggaaat	2520
ctttattaat	cacgtatggt	tcacagataa	ttcttttttt	aaaaaaaccc	aacctcctag	2580
agaagcacaa	ctgtcaagag	tcttgtacac	acaacttcag	ctttgcatca	cgagtcttgt	2640
attccaagaa	aatcaaaagt	gtacaatttg	tttgtttaca	ctatgatact	ttctaaataa	2700
actctttttt	tttaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaggggg	gggggggggn	2760
ggggccccc	cccccccaa					2780

&lt;210&gt; 88

&lt;211&gt; 1061

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens.

&lt;400&gt; 88

aattcggcac	gagagaagga	aatacatcaa	aatgcccaca	ttggttatct	gtagaggata	60
gaatgaaaga	tggttttatt	ctcttgtttg	ctcttattaa	agcaatcaga	tggtgcttct	120
cctgtactca	gagccctggc	tgcttctctg	ctggcctctc	ctgcgggctg	ctgtggaacc	180
agaaaagcct	taaacggaaa	tgtgggagag	aagggtggat	tcactttcat	gtctttccag	240
ggttgtgacc	cctcaagtcc	tggttgccct	tgctgttctc	tattaccttc	aaacagccag	300
ctcgtcttta	tttctttttt	agttttgtcg	gggttggtct	gatagatggt	agtccatcat	360
agccagatgt	gtctagcctt	gtcttttgaa	tgcaagattt	aggatgtggg	tacttagctg	420
ttagtggaca	tcagagtcac	tagtcaggat	gaaagagttc	ttggctttta	ctcccagaaa	480
ttctggtaac	gtcatgtata	gtgacggccg	catgtctaac	aggtggccag	gtaagtcttt	540
tggggtggtc	tgtgaatcac	agtttgggag	acattgactt	ttagggaggt	tgttctgaat	600
tcactagata	atagagatat	aatacagagc	tttgaaagct	ggtgtcttga	tgacagagcc	660
gtggcaatgg	ggagggttga	ggaggtggct	gttgggcctg	tctcctgggt	agagttgaaa	720
gggcctgaac	tcaagcagag	gcctcagaac	cgaaagggtg	tggaaggatg	cagcaagagg	780
cgccacacag	gagtactctg	cgccctggca	gggtctgaat	acacgtggga	gtgggtgagag	840
ggagaacttt	aagtccaggt	tttgtgcctc	agtgacttag	tgtggccata	tcattagaaa	900
tgtgttgagg	ccgggcacag	tggctcatgt	ctgtaatccc	agcactttga	gaggctgagg	960
caggaggatg	gcttgaggcc	aggagtttaa	aaccagcctg	gacaacatag	tgagagcctg	1020
tctctacaaa	aaaaaaaaaa	aaaaactcga	ggggggggccc	g		1061

&lt;210&gt; 89

&lt;211&gt; 1342

&lt;212&gt; DNA



&lt;213&gt; Homo sapiens

&lt;400&gt; 89

ccacgcgtcc	gggcgcgcgcg	gtgaaaggcg	cattgatgca	gcctgcggcg	gcctcggagc	60
gcgccggagc	agacgctgac	cacgttcctc	tctcgggtct	cctccgcctc	cagctccgcg	120
ctgcccggca	gcccggagcc	atgcgacccc	agggccccgc	cgctccccg	cagcggtcc	180
gcgccctcct	gctgctcctg	ctgctgcagc	tgcccgcgcc	gtcgagcgcc	tctgagatcc	240
ccaaggggaa	gcaaaaggcg	cagctccggc	agagggaggt	ggtggacctg	tataatggaa	300
tgtgcttaca	agggccagca	ggagtgcctg	gtcgagacgg	gagccctggg	gccaatggca	360
ttccgggtac	acctgggatc	ccaggtcggg	atggattcaa	aggagaaaag	ggggaatgtc	420
tgagggaaa	ctttgaggag	tcctggacac	ccaactacaa	gcagtgttca	tggagtccat	480
tgaattatgg	catagatctt	gggaaaattg	cggagtgtac	atttaciaag	atgcgttcaa	540
atagtgtctt	aagagttttg	ttcagtggct	cacttcggct	aaaatgcaga	aatgcagtct	600
gtgagcgttg	gtatttcaca	ttcaatggag	ctgaatgttc	aggacctctt	cccattgaag	660
ctataattta	tttggaccaa	ggaagccctg	aaatgaattc	aacaattaat	attcatcgca	720
cttcttctgt	ggaaggactt	tgtgaaggaa	ttggtgctgg	attagtggat	gttgctatct	780
gggttggcac	ttgttcagat	tacccaaaag	gagatgcttc	tactggatgg	aattcagttt	840
ctcgcatcat	tattgaagaa	ctaccaaaat	aaatgcttta	attttcatct	gctacctctt	900
tttttattat	gccttggaat	ggttcactta	aatgacattt	taaataagtt	tatgtatata	960
tctgaatgaa	aagcaaaagc	aaatatgttt	acagacccaa	gtgtgatttc	acactgtttt	1020
taaatctagc	attattcatt	ttgcttcaat	caaaagtggg	ttcaatatct	tttttagttg	1080
gttagaatac	tttcttcata	gtcacattct	ctcaacctat	aatttggaat	attgttggtg	1140
tcttttgttt	tttctcttag	tatagcattt	ttaaaaaat	ataaaagcta	ccaatctttg	1200
tacaatttgt	aaatgttaag	aatttttttt	atatctgtta	aataaaaaat	atttccaaaa	1260
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1320
aaaaaaaaaa	aaaaaaaaaa	aa				1342

&lt;210&gt; 90

&lt;211&gt; 770

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (690)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (762)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 90

ggtaggggtcg	ggkcgaccyr	cgcgtccggc	tcttcctcag	gcgccggcca	tggcgggaca	60
ggaggatccg	gtgcagcggg	agattcacca	ggactgggct	aaccgggagt	acattgagat	120
aatcaccagc	agcatcaaga	aaatcgcaga	ctttctcaac	tcgttcgata	tgtcttgctg	180
ttcaagactt	gcaacactaa	acgagaaaat	gacagccctt	gaacggagaa	tagagtacat	240
tgaagctcgg	gtgacaaaag	gtgagacact	cacctagaac	agtgcctgtc	tgctgctggg	300
aagttgcttt	acacaacaca	ggccacatgg	gaaaggcccc	agcagccttc	agctccttcc	360
tttctcctta	aagagcaaca	gggcttatcc	ttgtttttct	tttttcaaaa	gtgtggcctt	420
tgggctctgc	catctggggg	gtggtgtggt	atgtgggaag	aagttcagag	gaaccgttgg	480
aaacgacgtt	aggcatttta	ccttttcagt	aacattttat	acatctactt	gtcaatgtat	540
ttgagacatt	cacagccaaa	agcctgggac	tctttgtgaa	ggctcctctc	amctctatct	600
ttctttctct	ctctctcaaa	ctttccttaa	agttctcatt	gcctttgcac	tgcttctgtg	660
aacagtcttt	gtctcctccc	cacctttggn	ggaaagtgcg	gggcagtcct	ggtcaagaca	720
ctcatgcctt	ggcaatgtgg	cctgcagaga	atgttggtgt	anccaccagt		770

<210> 91  
 <211> 1570  
 <212> DNA  
 <213> Homo sapiens

<400> 91  
 gatggtttta ctgaagttgc taaaagttaa cagaaaagga agtgcaggaa catttcacaa 60  
 atctacaate tgtgagtatc acatcctgta tagctgtmaa cactggaata aggaagggct 120  
 gatgactttc agaagatgaa ggtaagtaga aaccgttgat gggactgaga aaccagagtt 180  
 aaaacctctt tggagcttct gaggactcag ctggaaccaa cgggcacagt tggcaacacc 240  
 atcatgacat cacaacctgt tcccaatgag accatcatag tgctcccatc aaatgtcatc 300  
 aactttctccc aagcagagaa acccgaaccc accaaccagg ggcaggatag cctgaagaaa 360  
 catctacacg cagaaatcaa agttattggg actatccaga tcttggtgtg catgatggta 420  
 ttgagcttgg ggatcatttt ggcactctgt tcttctctct caaattttac ccaagtgact 480  
 tctacactgt tgaactctgc ttaccatttc ataggaccct ttttttttat catctctggc 540  
 tctctatcaa tcgccacaga gaaaaggtrr accaagcttt tgggtgcatag cagcctgggt 600  
 ggaagcattc tgagtgtctc gtctgccttg gtgggtttca ttatcctgtc tgtcaaacag 660  
 gccaccttaa atcctgcctc actgcagtgt gatttgaca aaaataatat accaacaaga 720  
 agttatgttt cttactttta tcatgattca ctttatacca cggactgcta tacagccaaa 780  
 gccagtctgg ctggawctct ctctctgatg ctgatttgca ctctgctgga attctgccta 840  
 gctgtgctca ctgctgtgct gcggtggaaa caggcttact ctgacttccc tgggagtgtg 900  
 cttttcctgc ctcacagtta cattggtaat tctggcatgt cctcaaaaat gactcatgac 960  
 tgtggatatg aagaactatt gacttcttaa gaaaaaggg agaaatatta atcagaaagt 1020  
 tgattcttat gataatatgg aaaagttaac cattatagaa aagcaaagct tgagtttctt 1080  
 aaatgtgaagc ttttaagta atgaacatta aaaaaaacca ttatttctact gtcatttaag 1140  
 atatgtgttc attggggatc tcttgatttg cctgacattg acttcagcaa aagcacgggg 1200  
 ctgtaaatca ccatttacta gattagccaa atagtctgaa tttccagaaa acaaggcaga 1260  
 atgatcattc ccagaaacat tcccagaaa atgtttccca gaaaactaga cagmatgatc 1320  
 attcaatgga tcacagtga gcaaaggaca caacttttta ttgtaccctt taattgtcaa 1380  
 caggagttaa ctgatttgtt gtggtgctca gactttttta tacagggtgt agtggtttat 1440  
 cctatgtatt ttaactcatt agtgcataaa ggcaagcccc atataatgaa gtctcagggt 1500  
 atatgaaagt agctggcttc aaaataaaat ttttgagtgc aaaaaaaaaa aaaaataaaa 1560  
 aaaaaaaaaa 1570

<210> 92  
 <211> 2950  
 <212> DNA  
 <213> Homo sapiens

<400> 92  
 cccggctccc gccgcgtccc agccggggccc cccagcggtc ggcgggacgg ctcccggctg 60  
 cagtctgccc gccgcgcccg cgcgggggccc gagtcgcgaa gcgcgcctgc gaccggcggt 120  
 ccgggcgcgc tggagaggac gcgaggagcc atgaggcgcc agctgcgaag gtggcggcgc 180  
 tgctgctcgg gctgctcttg gactgcacag aagccaaaaa gcattgctgg tatttcgaag 240  
 gactctatcc aacctattat atatgccgt cctacgagga ctgctgtggc tccaggtgct 300  
 gtgtgcgggc cctctccata cagaggctgt ggtacttctg gttccttctg atgatggcg 360  
 tgcttttctg ctgcggagcc ggcttcttca tccggaggcg catgtacccc ccgcgctga 420  
 tcgaggagcc agccttcaat gtgtcctaca ccaggcagcc cccaaatccc ggcccaggag 480  
 cccagcagcc ggggcccggccc tattacacyg acccaggagg accggggatg aacctgtctg 540  
 ggaattccat ggcaatggct ttccagggtc caccacaact acccaggggg agtggtggcct 600  
 gccgcgccccc tccagctcac tgcaacacgc ctccgcccc gtacgaacag gtagtgaagg 660  
 ccaagtagtg ggggtgcccac gtgcaagagg agagacagga gagggccttt cctggcctt 720  
 tctgtcttct ttgatgttca cttccaggaa cgggtctcgt ggctgctaag ggcagttcct 780  
 ctgatatact cacagcaagc acagctctct ttcaggcttt ccatggagta caatatatga 840  
 actcacactt tgtctctctt gttgtctctg tttctgacgc atctgtgctc tcacatggta 900  
 gtgtggtgac agtccccgag ggctgacgtc cttacgggtg cgtgaccaga tctacaggag 960  
 agagactgag aggaagaagg cagtgtctga ggtgcagggt gcatgtagag gggccaggcc 1020  
 gagcatccca ggcaagcatt cttctgcccc ggtattaata ggaagcccca tgccggggcg 1080

ctcagccgat	gaagcagcag	ccgactgagc	tgagcccagc	aggatcatctg	ctccagcctg	1140
tcctctcgtc	agccttcctc	ttccagaagc	tggtggagag	acattcagga	gagagcaagc	1200
cccttgatcat	gtttctgtct	ctgttcatat	cctaaagata	gacttctcct	gcaccgccag	1260
gaaagggtag	cacgtgcagc	tctcaccgca	gatggggcct	agaatcaggc	ttgcttggag	1320
gcctgacagt	gatctgacat	ccactaagca	aatttattta	aattcatggg	aaatcacttc	1380
ctgccccaaa	ctgagacatt	gcattttgtg	agctccttgg	ctgatttggg	gaaaggactg	1440
ttaceccattt	ttttggtgtg	tttatggaag	tgcatgtaga	gcgtcctgcc	ctttgaaatc	1500
agactgggtg	tgtgtcttcc	ctggacatca	ctgcctctcc	agggcattct	caggccccggg	1560
ggtctccttc	cctcaggcag	ctccagtggg	gggttctgaa	gggtgctttc	aaaacggggc	1620
acatctggct	gggaagtcac	atggactctt	ccaggggagag	agaccagctg	aggcgtctct	1680
ctctgaggtt	gtgttgggtc	taagcgggtg	tgtgtctggg	tccaaggagg	aggagcttgc	1740
tgggaaaaga	cagggaagt	actgactcaa	ctgactgac	catgttgtca	taattagaat	1800
aaagaagaag	tggtcggaaa	tgacatttcc	tggataggaa	tcacagctca	ccccaggatc	1860
tcacaggtag	tctcctgagt	agttgacggc	tagcggggag	ctagtctccg	cgcatagtta	1920
tagtgttgat	gtgtgaacgc	tgacctgtcc	tgtgtgctaa	gagctatgca	gcttagctga	1980
ggcgcctaga	ttactagatg	tgctgtatca	cggggaatga	ggtgggggtg	cttatttttt	2040
aatgaactaa	tcagagcctc	ttgagaaatt	gttactcatt	gaactggagc	atcaagacat	2100
ctcatggaag	tggatacggg	gtgatttggg	gtccatgctt	ttcaactctga	ggacatttaa	2160
tcggagaacc	tcctggggaa	ttttgtggga	gacacttggg	aacaaaacag	acaccctggg	2220
aatgcagttg	caagcacaga	tgctgccacc	agtgtctctg	accaccctgg	tgtgactgct	2280
gactgccagc	gtggtacctc	ccatgctgca	ggcctccatc	taaatgagac	aacaaagcac	2340
aatgttcaact	gtttacaacc	aagacaactg	cgtgggtcca	aacactcctc	ttcctccagg	2400
tcatttgttt	tgcatTTTTA	atgtctttat	tttttgaat	gaaaaagcac	actaagctgc	2460
ccctggaatc	gggtgcagct	gaataggcac	ccaaaagtcc	gtgactaaat	ttcgtttgtc	2520
tttttgatag	caaattatgt	taagagacag	tgatggctag	ggctcaacaa	ttttgtattc	2580
ccatgtttgt	gtgagacaga	gtttgttttc	ccttgaactt	ggttagaatt	gtgctactgt	2640
gaacgctgat	cctgcatatg	gaagtccrc	ttcgggtgaca	tttcttggcc	attcttgttt	2700
ccattgtgtg	gatgggtggg	tgtgcccact	tcctggagtg	agacagctcc	tggtgtgtag	2760
aattccccga	gcgtccgtgg	ttcagagtaa	acttgaagca	gatctgtgca	tgcttttctc	2820
ctgcaacaat	tggctcgttt	ctcttttttg	ttctcttttg	ataggatcct	gtttcctatg	2880
tgtgcaaaat	aaaaataaat	ttggggcaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	2940
aaaaaaaaag						2950

&lt;210&gt; 93

&lt;211&gt; 1722

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 93

ggcacgagcc	agagcaggct	gctaggcctg	gggccaccac	tgcccctggg	tgctacaccc	60
agtgtgctgg	gtcactggga	acttcttgaa	gtggtgtcac	ctgaactggg	cccccaagga	120
tggggtgcgg	gcagtaccgc	aggaagagga	gcagcccctg	tgaagattga	gagctgccag	180
aggctctgtg	attggctgcg	gcacgatgac	ccgcgcacgg	attggctgct	tcgggcccggg	240
gggcccggcc	cgggggacag	aatccgcccc	cgaaccttca	aagagggtac	cccccggcag	300
gagctggcag	accyaggagg	tgcgacagac	ccgcgggggc	aacggactgg	ggccaagagc	360
cgggagcgcg	ggcgcaaagg	caccagggcc	cgcccaggcc	gccgcgcagc	acggccttgg	420
gggttctgcg	ggccttcggg	tgcgcgcttc	gcctctagcc	atgggggtccg	cagcgttggg	480
gatcctgggc	ctggtgctgt	gcctgggtggg	ctgggggggt	ctgacctctg	cgtgcgggct	540
gccccatgtg	caggtgaccg	ccttcctgga	ccacaacatc	gtgacggcgc	agaccacctg	600
gaaggggctg	tggatgtcgt	gcgtgggtgca	gagcacsggg	cacatgcagt	gcaaagtgtg	660
cgactcgggtg	ctggctctga	gcaccgaggt	gcaggcggcg	cgggcgctca	ccgtgagcgc	720
cgtgctgctg	gcgttcgttg	cgctcttcgt	gaccttggcg	ggcgcgagc	gcaccacctg	780
cgtggccccg	ggcccggcca	argcgctgtg	ggccctcacg	ggaggcgtgc	tctacctgtt	840
ttgcgggctg	ctggcgctcg	tgccactctg	ctgggttcgc	aacattgtcg	tcgcgagatt	900
ttacgacccg	tctgtgcccc	tgtcgcagaa	gtacgagctg	ggcgagcgc	tgtacatcgg	960
ctgggcccgc	accgcgctgc	tcattgtagg	cggctgcctc	ttgtgctgcg	gcgcctgggt	1020
ctgcaccggc	cgccccgacc	tcagcttccc	cgtgaagtac	tcagcgccgc	ggcgggccac	1080
ggccaccggc	gactacgaca	agaagaacta	cgtctgaggg	cgctgggcac	ggccggggcc	1140

ctcctgccag	ccacgcctgc	gaggcggttg	ataagcctgg	ggagccccgc	atggaccgcg	1200
gcttccgccc	ggtagcgccg	cgcgaggct	cctcggaacg	tccggctctg	cgccccgacg	1260
cggtccctgg	atccgctcct	gcctgcgccc	gcagctgacc	ttctcctgcc	actagcccgg	1320
ccctgccctt	aacagacgga	atgaagtttc	cttttctgtg	cgcgcgctg	tttccatagg	1380
cagagcggtg	gtcagactga	ggatttcgct	tccccctcaa	gacgctgggg	gtcttggtg	1440
ctgccttact	tcccagaggc	tcctgctgac	ttcggagggg	cgatgcaga	gcccagggcc	1500
cccaccggaa	gatgtgtaca	gctggctctt	actccatcgg	cagggccga	gcccagggac	1560
cagtgccttg	gcctggacct	cccgtctca	ctccagcatc	tcccaggca	aggcttgtgg	1620
gcaccggagc	ttgagagagg	gcgggagtgg	gaaggctaag	aatctgctta	gtaaatggtt	1680
tgaactctca	aaaaaaaaa	aaaaaaaaa	aaaaaaaaa	aa		1722

<210> 94  
 <211> 635  
 <212> DNA  
 <213> Homo sapiens

<400> 94						
gcctaagag	agctccccca	ggaccagccc	tggccaaggg	attgctgcag	ccctcatcca	60
cttccaagc	actggaaaca	aacattggag	accaagttag	gcgtcactca	acagccgtag	120
taatcaggg	aatgacaagt	tacatactga	tatcctttgt	tttgcgtgatt	ggagttgggt	180
gcattgaaa	agatcagtcg	tgcccagtg	ttgggggaag	gaagcgtctt	cacctgttgt	240
ttgtgggag	acagttgagg	caggtgagga	tgctgagagg	tgagctcagc	tgtgcctggt	300
accgtccaca	tgtgcaagcc	cttcagctcg	gtggtgtgtac	ttgtttttga	gatgcagttt	360
cactcttgtc	acccaggctg	gagtgcatgg	catgatcttg	gctcgtgca	acatccgcct	420
cccgggttca	agcgattctc	ctgtctctac	taaaaatata	aaaattagct	gggtgtgggt	480
gtgcgtgect	ttaatcccag	ctactcagaa	ggctgaggtg	caagaattgc	ttgaacctgg	540
gaggtggagg	ttgccgtggg	ccgagatcac	gccaccgcac	tccagcctag	gcaacagagc	600
tagactgtct	caaaaaaaaa	aaaatgacct	tcgag			635

<210> 95  
 <211> 3798  
 <212> DNA  
 <213> Homo sapiens

<400> 95						
ccacgcgtcc	ggggcttcat	acaggaaatc	tattgctgtg	tcaagttcca	gagaaaagct	60
tctgttcgtc	caagttacta	accaggctaa	accacataga	cgtgaaggaa	ggggctagaa	120
ggaaggaggt	gcccactgt	tgatggggta	agaggatcct	gtactgagaa	gttgaccaga	180
gagggctca	ccatgcgcac	agttccttct	gtaccagtg	ggaggaaaag	tactgagtga	240
agggcagaaa	aagagaaaac	agaaatgctc	tgcccttggg	gaactgctaa	cctagggcta	300
ctgttgattt	tgactatctt	cttagtggtc	gaagcggagg	gtgctgctca	accaaacaac	360
tcattaatgc	tgcaaaactag	caaggagaat	catgctttag	cttcaagcag	tttatgtatg	420
gatgaaaaac	agattacaca	gaactactcg	aaagtactcg	cagaagttaa	cacttcattg	480
cctgtaaaga	tggctacaaa	tgctgtgctt	tgttgccctc	ctatcgcatt	aagaaatttg	540
atcataataa	catgggaaat	aatcctgaga	ggccagcctt	cctgcacaaa	agcctacaag	600
aaagaaacaa	atgagaccaa	ggaaaccaac	tgtactgatg	agagaataac	ctgggtctcc	660
agacctgac	agaattcgga	ccttcagatt	cgtaccgtgg	ccatcactca	tgacgggtat	720
tacagatgca	taatggtaac	acctgatggg	aatttccatc	gtggatatca	cctccaagtg	780
ttagttacac	ctgaagtgac	cctgtttcaa	aacaggaata	gaactgcagt	atgcaaggca	840
gttgccagga	agccagctgc	gcatactctc	tggatcccag	agggcgattg	tgccactaag	900
caagaatact	ggagcaatgg	cacagtgcct	gttaagagta	catgccactg	ggaggtccac	960
aatgtgtcta	ccgtgaactg	ccacgtctcc	catgtgactg	gcaacaagag	tctgtacata	1020
gagctacttc	ctgttccagg	tgccaaaaaa	tcagcaaaat	tatatattcc	atatatcatc	1080
cttactatta	ttattttgac	catcggtggg	ttcatttggg	tgttgaaagt	caatggctgc	1140
agaaaatata	aattgaataa	aacagaatct	actccagttg	ttgaggagga	tgaaatgcag	1200
ccctatgcca	gctacacaga	gaagaacaat	cctctctatg	atactacaaa	caaggtgaag	1260
gcactctgagg	cattacaaaag	tgaagttgac	acagacctcc	atactttata	agttgttggg	1320

ctctagtacc	aagaaacaac	aacaaacgag	atacattata	attactgtct	gattttctta	1380
cagttctaga	atgaagactt	atattgaaat	taggttttcc	aagggtctta	gaagacattt	1440
taatggattc	tcattcatac	ccttgtataa	ttggaatttt	tgattccttag	ctgctaccag	1500
ctagttctct	gaagaactga	tgttattaca	aagaaaatac	atgcccataga	ccaaatatct	1560
aaattgtgca	ggacagtata	taatgaaaac	caaattttcct	caagaaataa	ctgaagaagg	1620
agcaagtgtg	aacagtttct	tgtgtatcct	ttcagaatat	tttaattgtac	atatgacatg	1680
tgtatatgcc	tatggtatat	gtgtcaattt	atgtgtcccc	ttacatatatac	atgcacatat	1740
ctttgtcaag	gcaccagtgg	gaacaatata	ctgcattact	gttctatata	tatgaaaacc	1800
taataatata	agtcttagag	atcatttttat	atcatgacaa	gtagagctac	ctcattcttt	1860
ttaatgggta	tataaaattc	cattgtatat	ttatatcatt	atttaattaa	aaacaaccct	1920
aatgatggat	atthagattc	ttttaagttt	tgttttatttc	ttttaagttt	tgtttgtggg	1980
ataaacaata	ccacatagaa	tgtttcttgt	gcatatatct	ctttgttttt	gagtatatct	2040
gtaggataac	tttcttgagt	ggaattgtca	ggtcaaaggg	tttgtgcatt	ttactattga	2100
tatatatggt	aaattgtgtc	aaatatatat	gtcaaattcc	ctccaacatt	gtttaaatgt	2160
gcctttccct	aaatttctat	tttaataact	gtactattcc	tgcttctaca	gttgccaatt	2220
tctcttttta	atcaaccaga	ttaaatatga	tgtgagatta	taataagaat	tatactattt	2280
aataaaaaatg	gattttatatt	tttggctatg	tttgaagag	agtgaatgca	cgtgtgagaa	2340
cattagcttc	ttctgaactc	attatatctc	cacagagggtg	ttgatacttg	atgcctaaca	2400
gttttgcaga	tgtgctacat	tggaattgtg	tatttttatg	gtgtacattc	tattgtgata	2460
tattttattga	ataattaatg	tctattgacc	atataagtgg	cgaaaaatgc	accatagagg	2520
acatggggta	tttattttaca	aactatgagc	tacataataa	gcaagtggcc	atgggatggc	2580
atgaccctcc	cctccatatt	tttgtggagc	aaaatattgg	caatgtttat	gtaaatcatt	2640
gttaatatca	tgaaattatt	tttaattaaa	aacataagtc	tatttgctcc	atagcagaaa	2700
aaacatgaga	agttttttca	tcatgataga	aattgaaaca	aaatatattc	attcttcaat	2760
cataccatct	gagattttta	agacagctat	tttgtcttat	aagtatatatt	ttctccctct	2820
agacatttca	gttactatgg	attttgtcct	caaagggacc	tttagtctat	tttggatgta	2880
aagctaattct	aatgacactt	ggcacatgat	attttgatca	agccattttg	acttgaccaa	2940
aaagcagtgt	ccattaggtt	tctgcatata	aatattacca	agcaatgttc	acaatagaca	3000
tcattacact	gtccttgaaa	tttattaatt	cttcatccaa	ccctgggtga	gctgaggctc	3060
atagttaggt	tcaagactat	ctgtttaaat	attactgaaa	aacaaagtaa	gacagtacta	3120
tgcttacctc	ttaacttgat	aatgtcaaaa	caggcatgtt	aatgacatc	atagaaaaga	3180
cttcaagata	atttatagaa	gttaaattat	attgtacaga	aaataattgt	atgaaaatct	3240
ctactatggg	gctggaacat	ggttgaacat	tagaatgata	taaaaaatta	tatatattct	3300
ccaaatccac	gctagacctg	tcaaattaga	gaatctagag	attagacctg	gcgtgtcagc	3360
aaggctcatcc	aggaagcaga	ggctgagacg	gagttaggtg	tgattactta	catagtcgat	3420
tacatttttac	aaataacatt	ttatatgtct	cattttactgt	gctttctccc	catcccatct	3480
tgtatctttt	cctttgtctt	gctagatttg	tcaattttct	ctctctttct	gtctctctct	3540
ctttcaatat	ctctaataat	ttgaaagtaa	ttcatcataa	ctaaatatct	attgggggta	3600
tgcttcactt	acaaacttct	gaaaacggct	ttactgagat	ataattgata	tatttaagtg	3660
tacagtttgt	taaattttgc	acatatttaa	aatgtggact	ttggtaaatg	ttgacatagt	3720
tttacatctg	tgaaccatc	agcataatca	agataataaa	cttgtccatc	accccccaaa	3780
aaaaaaaaaa	aaaaaaaaa					3798

&lt;210&gt; 96

&lt;211&gt; 2683

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (2640)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (2676)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 96

acaasgtmac	gcctgacagg	tmaccggatc	cggggaattcc	cggtgcgacc	cacggcgctcc	60
gcattttgcaa	taacagaaaa	ggaattgcat	gtatgaagtt	ttcaatcgtg	ggctttttctt	120
tgttgtgggg	aggggggtcgg	gggatagttt	gatttccatt	ttctgaaaac	gacagacttg	180
gattctgttt	gtgtgtgcat	attttatcca	gccttaagtt	ataaagctca	tctgtcccgc	240
tgcatccct	gtgtattttc	aggacatggc	tcgtgggtgt	gtgtgttcat	tgtgtgcgtc	300
tgtatgtatt	tttctgtcat	cactgttccc	tctcctcccg	agtgtgcatt	cagttaatat	360
aatcagttgc	ttgcttcttt	caaagtgcct	tgaagggtct	gaactcatgt	gtgagcatct	420
ttatcaacta	tcccaattgc	atgttctcca	tcacatattc	tcttatttgc	tctgtacccc	480
ctgagaatat	gttttagaga	tattggaata	aagctgtctg	ggtaaggagt	aggcttagcc	540
gacctatgaa	taatacactt	tagtctagtt	ctttattcta	aatctggatt	gccagttattg	600
tgtattttaa	ccaagtctgt	gaatacctgc	tttttttggc	cacagagtaa	caagttttca	660
tgtaagatct	tcatacaaa	gtaggaagta	aaaatagctt	agaaagctct	gtcaggtggt	720
ttgtgcagct	gacagargta	atgttacatc	acctaaaaaa	gaaagatata	cggtcagtta	780
tcctaaaaat	aaattgtttg	gaaagtacaa	tgcaccacat	ttttgtagaa	gtctactatt	840
tgataaacag	ttgaaattca	agatgtgttt	gaccttagt	catttttact	ctttggttct	900
gagtatacct	attttcttag	cgtatctgcc	ttgtttatct	ttttcttcac	cttttaacaa	960
gtatgacata	ggaaagtcat	tttttttag	aattcatgga	tcagtctgat	ctactcttat	1020
tcataatgga	acatgtaaat	atactgaaaa	ctgtttttca	ggagagaaat	atgagttgga	1080
gggaaggaaa	agtggttcta	ctaattgttc	aaaatcctca	tcagagaagg	tatgatgttc	1140
tcaggtgtgg	aaaatatttt	ttagttgatt	gagaatgcag	gtttaacaga	agagataagg	1200
ggcataatga	ctgctgggtt	tccagactgg	attttcttac	cgcaactatt	aatgttctca	1260
gagttgatga	ggaccacctt	tgtgtatata	cttgtagttt	taaaccttgc	attggtaaca	1320
aaatgatcaa	ctttaatcca	ggtagaattc	aagatggctg	tacttcagtt	gtatgataaa	1380
attaatgggt	ctcatgactt	gtgtggcatc	taaaaataat	gtttttatag	catctctctg	1440
ccactaaatt	gttgacttga	atttttggaa	aaaaaaaagt	tggtgttgat	atgtatatgt	1500
gtgtgtgtat	atatgtattt	ataaacaagt	gtgtttgagt	aacaagtgag	tttcatagtc	1560
ttcccctacg	catgtgtatt	ccacacacaa	atggctgagt	tatagtcata	aaacaatttg	1620
caataaaaaa	aaaaccacaa	cagattgtca	gttaaccagg	aaacagttaa	tgttttttaa	1680
tgaatctggc	attatagtga	gcaaattgctg	tattaattta	ggctaatttc	taatactacc	1740
ataattttgtg	tctaaatttc	tggtggggta	gaaattacta	aaattgtggg	gagtttttct	1800
tgattttttac	attgcttttag	gaaacatttt	tactaattca	gctgtcttag	gtaaaaatgaa	1860
tagttttctt	cctgtttttt	tatgtgtcat	tgtttagtgg	ctcagaattc	tgatcagtaa	1920
ctttgtgtat	gatgtggaat	tacaaaccgt	ttgaatgac	cagttgaaaa	cgtatccctc	1980
tactttcttc	agttgtagaa	aagggttaatt	tccctcagtg	tcccacatta	taccaaccta	2040
agagaagaac	aggtaatagg	gagaaataaa	catacgggtg	tttcagtggt	tttgggtcatg	2100
tgctccacag	agaaactaac	cattcagttg	tcttaatttt	agttcgttct	acctgtgag	2160
gagtttgttt	ccatcagttg	ttgactttcc	aaaatgttgc	attaagtaat	agttgtcact	2220
ctgttgggtct	catgggtcaat	atcaatcaga	ctttcatgat	ctctactaat	tattagtaga	2280
gtcctgtact	atgtctgtaa	ctactaagtt	taaagaaaag	cacatagtca	cttcatctct	2340
ttttttctta	gcctacgtc	actccccaac	ccatcccaac	attgacatgc	tatctgtgga	2400
caaatagcag	ttctcagaat	ctagtcaagt	tgccatcatc	ccccttgcc	tggccgttca	2460
tagtaggtat	gcatatgttt	gtttctgtac	agtactgtgt	gtgtgtgtgt	atatatatat	2520
acatctgtat	gcacacatct	ttgataaaat	agctatttga	ctagcagggt	taaagtggct	2580
tttaattact	tcgtgagtgt	tattggatac	atcttaaaaa	aaaaaaatct	ggaccacagan	2640
ccatgccata	cttgggttga	ctattttggg	gcattnaaaa	ttg		2683

&lt;210&gt; 97

&lt;211&gt; 2181

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (5)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 97

gtacnggatt	cccgggtcga	cccacgertc	cggcaggcag	aggggcagtg	ggcgcgggga	60
agatgattct	gggcctcccc	catctactgt	cattaaccaa	aatgaaacat	ttgccâacat	120
aattttttaa	cctactgtag	tacaacaagc	caggattgcc	cagaatggaa	ttttgggaga	180
ctttatcatt	agatatgacg	tcaatagaga	acagagcatt	ggggacatcc	aggttctaaa	240
tggctatttt	gtgcactact	ttgctcctaa	agaccttcct	cctttaccca	agaatgtggt	300
attcgtgctt	gacagcagtg	cttctatggt	gggaaccaa	ctccggcaga	ccaaggatgc	360
cctcttcaca	attctccatg	acctccgacc	ccaggaccgt	ttcagtatca	ttggattttc	420
caaccggatc	aaagtatgga	aggaccactt	gatatcagtc	actccagaca	gcatcaggga	480
tgggaaagtg	tacattcacc	atatgtcacc	cactggaggc	acagacatca	acgggggtcct	540
gcagagggcc	atcaggctcc	tcaacaagta	cgtggccac	agtggcattg	gagaccggag	600
cgtgtccctc	atcgtcttcc	tgacggatgg	gaagccccac	gtcggggaga	cgcacacctt	660
caagatcctc	aacaacaccc	gagaggccgc	ccgaggccaa	gtctgcatct	tcaccattgg	720
catcggcaac	gacgtggaçt	tcaggctgct	ggagaactg	tcgctggaga	actgcccctt	780
cacacggcgc	gtgcacgagg	aggaggacgc	aggctcgag	ctcatcggtt	tctacgatga	840
aatcaggacc	ccgctcctct	ctgacatccg	catcgattat	ccccccagct	cagtgggtgca	900
ggccaccaag	accctgttcc	ccaactactt	caacggctcg	gagatcatca	ttgcggggaa	960
tgggtgtgac	aggaagctgg	atcacctgca	cgtggaggct	accgccagca	acagtaagaa	1020
attcatcate	ctgaagacag	atgtgcctgt	gcgccctcag	aaggcaggga	aagatgtcac	1080
aggaagcccc	aggcctggag	gcgatggaga	gggggacmcc	aaccacatcg	agcgtctctg	1140
gagctacctc	accacaaagg	agctgctgag	ctcctggctg	caaagtgcag	atgaaccgga	1200
gaaggagcgg	ctgcggcagc	ggggccaggc	cctggctgtg	agctaccgct	tcctcactcc	1260
cttcacctcc	atgaagctga	gggggcccgt	cccacgcag	gacggcctgg	aggaggccca	1320
cggcatgtcg	gctgccatgg	gacccgaacc	ggtggtgcag	agcgtgcgag	gagctggcac	1380
gcagccagga	cctttgtctca	agaagccata	ccagccaaga	attaaaatct	ctaaaacatc	1440
agtggatggt	gatccccact	ttgttgtgga	tttccccctg	agcagactca	ccgtgtgctt	1500
caacattgat	gggcagcccc	gggacatcct	caggctggct	tctgatcaca	gggactctgg	1560
tgtcacagtg	aacggagagt	taattggggc	acccgcccct	ccaaatggcc	acaagaaaca	1620
gcgcacttac	ttgcgcacta	tcaccatcct	catcaacaag	ccagagagat	cttatctcga	1680
gatcacaccg	agcagagtca	tcttgatgg	tggggacaga	ctgggtgctc	cctgcaacca	1740
gagtgtggtg	gtggggagct	ggggkctgga	ggtgtccgtg	tctgccaacg	ccaatgtcac	1800
cgtcaccatc	cagggtctca	tagcctttgt	catcctcatc	cacctctaca	aaaagccggc	1860
gcccttccag	cgacaccacc	tgggtttcta	cattgccaac	agcgagggcc	tttccagcaa	1920
ctgccacgga	ctgctgggtc	agttcctgaa	tcaggatgcc	agactcacag	aagaccctgc	1980
agggcccagc	cagaacctca	ctcaccctct	gctccttcag	gtgggagagg	ggcctgaggc	2040
cgtcctaaca	gtgaaaggcc	accaagtccc	agtggctctg	aagcaaagga	agatttacaa	2100
cggggaagag	cagwtagayt	gytggtttgc	caggaacatg	ccgccaact	gattgacggg	2160
gagtacagga	ttacctggca	t				2181

&lt;210&gt; 98

&lt;211&gt; 1957

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 98

acgcgtccgg	aggcggcagg	ctcatggcgc	cggcgtcgcg	gttgcctcg	ctctgggcgc	60
tggcggtctg	ggctctaccc	ggctccgggg	cggagggcga	cggcggtggg	cgcccgggcg	120
ggccgggggg	cgtggcggag	gaggagcgt	gcacggtgga	gcgtcgggcc	gacctcacct	180
acgcggagtt	cgtgcagcag	tacgccttcg	tcaggcccgt	catcctgcag	ggactcacgg	240
acaactcgag	gttccggggc	ctgtgctccc	gcgacaggtt	gctggcttcg	tttggggaga	300
gagtgttccg	gctgagcacc	gccaaacact	actcctacca	caaagtggac	ttgcccttcc	360
aggagtattg	ggagcagctg	ctgcaccccc	aggaccccac	ctccctgggc	aatgacaccc	420
tgtactttct	cggggacaac	aacttcaccg	agtgggcctc	tctctttcgg	cactactccc	480
cacccccatt	tggcctgctg	ggaaccgctc	cagcttacag	ctttggaatc	gcaggagctg	540
gctcgggggt	gcccttccac	tggcatggac	ccgggtactc	agaagtgatc	tacggctcgta	600
agcgtggtt	cctttaccca	cctgagaaga	cgccagagtt	ccacccaac	aagaccacgc	660
tggcctggct	ccgggacaca	tacccagcct	gcaccgtctg	cacggccctg	gagtgtacca	720
tccgggctgg	tgaggtgctg	acttcccagc	cgctgggtgg	atgctacgct	caaccttgac	780
accagcgtct	tcattctacc	ttcctcggtt	agccaaaaca	gctggcagga	ctgccggtca	840

cacaccagca	cgtccccacct	cgtgctcacg	gattttatta	cacagatagt	ggcggcaatg	900
gcctcagccc	agccccacct	cacctgcttt	tccagcccac	aaagggggac	gatcacggcc	960
cagcaaaaagc	gatgctgaga	ggggaaacag	tccagagtcc	aacagcagaa	cttgggggaa	1020
gcggtcgggg	tggccaggaa	cataaactat	gtataggggc	cgggggcttc	tgcccagggc	1080
tcccctggac	caggacgcca	ggtagggcag	ggaacctcag	tagtcctcca	cccagccatt	1140
ctcagagatg	aatgcgtcaa	taacctcctt	catagccaag	ttggggatga	gctgttcctg	1200
ggtcaggggg	ctccgggtca	cgggggtcaaa	atgacccaca	cgctgcagtg	acaagaaggg	1260
cagagggcag	tcatggggcc	caggaccatg	ccactggccc	tgctccccc	gccgcaggcc	1320
tcacctgcag	gtgctcctcg	atgtccttgc	ggtcgtaggt	gatgccactg	ggcgtgatgc	1380
acggctcccc	catcagctca	aagctgatct	tgccacacag	gtagtcgggg	atgtctcgct	1440
tctgtggcac	aggggcacac	ggtcagaggc	tgaaggggg	cactgcacga	gcacctgcca	1500
gccatcggca	gcaagcgaca	cacactcacc	ttcctcttct	catccacctg	agaaaaaagc	1560
tcgtccatgt	ccgccatgta	cttgtcctgt	gaagagttga	gtgctgtgct	tgggggagac	1620
accccacctc	cctcctccat	ggggcacaga	cccaacacaa	ggcggggatg	ctcccacgcc	1680
acgtgcacac	acacagaccc	acatgtgggt	ggggggcacc	ctcacgtgct	tggcctcaat	1740
gcaggcctgc	tgggcccggg	cgtggctgtc	gtcctcatca	ccctcgtggt	ttcgctggca	1800
ctcttccagc	tccctggggg	ttgaccagga	gccggtcaga	gatggacctg	gccagatgtc	1860
tgaccacacc	ccaatctcag	agctaacatc	cacacttccc	cacatttctc	gcttgccagt	1920
aaagccttcg	ataaacaana	aaaaaaaaaa	aaaaaaa			1957

&lt;210&gt; 99

&lt;211&gt; 1112

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 99

ccacgcgtcc	gcccaccaac	aaaggcactt	gcagttacat	tcaccacatt	tgtaacggag	60
ccattgaagc	atattggaaa	aggaactggg	gaatttatta	aagcactcat	gaaggaaatt	120
ccagcgctgc	ttcatcttcc	agtgtgata	attatggcat	tagccatcct	gagtttctgc	180
tatggtgctg	gaaaatcagt	tcattgtgctg	agacatatag	gcggtcctga	gagagaacct	240
cccaggcac	ttcgccacg	ggatagaaga	cggcaggagg	aaattgatta	tagacctgat	300
ggtggagcag	gtgatgccga	tttccattat	aggggccaaa	tgggccccac	tgagcaaggc	360
ccttatgcc	aaacgtatga	gggtagaaga	gagattttga	gagagagaga	tgttgacttg	420
agatttcaga	ctggcaacaa	gagccctgaa	gtgctccggg	catttgatgt	accagacgca	480
gaggcacgag	agcatcccac	ggtggtaccc	agtcataaat	cacctgtttt	ggatacaaa	540
cccaaggaga	caggtggaat	cctgggggaa	ggcacaccga	aagaaagcag	tactgaaagc	600
agccagtcgg	ccaagcctgt	ctctggccaa	gacacatcag	ggaatacaga	aggttcaccc	660
gcagcggaaa	aggccagct	caagtctgaa	gcgcaggca	gccagacca	aggcagaca	720
tacagccccg	caagaggtgt	ggctggacca	cgtggacagg	atccggtcag	cagccccctgt	780
ggctagagga	acaccagcac	aaacgacagc	ctcaagtctc	cttcgagctt	tgtaaagtaa	840
tctcacaagc	tgccaaagaa	cacatcccaa	gcagtaatca	cagatgatga	agcaccttac	900
aggacccctc	caccctcaaa	cgcgcatgtc	cagagaagta	ttaatgcctt	aatagactac	960
tgaagggttaa	ctattttacat	catttctaaaa	atatttttaag	ctgtattaca	gtgctataaaa	1020
ttctcctttt	aaggaaaaaa	gtaatatagatt	gtttccttac	cgtgtttgaa	aggaaaaaaa	1080
ataaatccat	agtaactaaa	aaaaaaaaaa	aa			1112

&lt;210&gt; 100

&lt;211&gt; 887

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (303)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 100



ggcacgagag	aattataggt	gcatgatggg	gcttttggag	actggcaatg	ttctgttttg	60
ggtctgggta	gtggttactt	gtgtgtattc	actttatgct	aactcattga	actgtacaga	120
tatggactgt	gcccccttct	atatgtgtgt	aatgcttcaa	caaaagtgtc	aatagtgtatg	180
catgcacatg	ttaaaatttc	aaactatatt	aaagagtgtg	caattaaaag	gaagttatcc	240
tctcactcta	aagtcctatt	tcctctcttc	agtctatcat	tactactagt	ttctagtata	300
tcnttttagaa	atgkgctgta	aaagaacaat	gatgtgtatg	tctctacagg	tatatattta	360
ttctttttgt	aattacaaaa	atataaaactc	actacatatg	ttattctacc	acatgctttt	420
tttcatgtaa	cagtatgtct	tagacatctt	ttcctattat	tgcgtggaag	ctatcaacct	480
tattctttgt	aatgactcca	taaattattc	tactgtgaaa	acacaccatg	tttaaccagt	540
tctctgttag	tgaacattta	ggatttttcc	agttttttaa	tattacagtg	acataattaa	600
cattaaacat	atcttttgac	acatgtcctt	gcacacatgt	ataggtatga	tggactttta	660
cacccttttg	ctagattctt	tagcacataa	cgtaaatatc	ccatagagtc	aaaaccaccc	720
ttaaaacttc	ctcaggaggg	tgggtgtggt	ggctcacggc	cgtaatccca	gcacttttag	780
aggccgaggt	gggcggatca	cgaggtcagg	agatcgagac	cacctgggcc	aacatggtga	840
aacctgtct	ctactaaaaa	tacaaaaaaa	aaaaaaaaaa	actcgta		887

&lt;210&gt; 101

&lt;211&gt; 1248

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 101

ccacgcgtcc	ggtctattaa	ggtgacagaa	ctgaagggcc	tggccaacca	tgtggttgtg	60
ggcagtgtct	cctgtgagac	caaggacctg	tttgccggcc	tgccccaggt	tgtggctgtg	120
gatatcaatg	accttgggtac	catcaagctc	agcctggaag	tcacatggag	ccccttcgac	180
aaggatgacc	agccctcagc	tgtttcttct	gtcaacaagg	cctccacagt	caccaagcgc	240
ttctccacct	atagccagag	cccaccggac	acacctcac	ttcggaaca	ggctttctat	300
aacatgctgc	gacggcagga	ggagctggag	aatgggacag	catggtccct	gtcatctgaa	360
tcttcagacg	actcatccag	cccacagctc	tcaggcactg	cccggcactc	accagcccct	420
aggcccctgg	tgcagcagcc	cgagcccctt	cccattccaag	ttgccttccg	caggcctgag	480
accccagct	ctgggcccct	ggatgaggag	ggggccgtgg	ccccagtcct	ggcaaatggg	540
catgcaccct	acagtccggac	tctgagccac	atcagtgagg	ctagtgtaaa	tgctgccttg	600
gctgaggctt	cagtggaggc	cgttggccca	aaaagcctat	cctggggacc	tagccccact	660
acacacccag	ctcccaccca	tggaaagcac	cccagtcctg	ttcctcctgc	cctggaccct	720
ggccactctg	ccacaagctc	taccctcggt	acaacaggct	ctgtccccac	atctacagac	780
cctgccccat	ctgcacacct	agactcagtt	cataagtcca	cagactctgg	cccttcagaa	840
ctgccaggcc	ccactcacac	cactacaggc	tctacctata	gtgccattac	cactacccac	900
agtgtccaa	gccccctcac	tcacactact	acaggctcca	cccacaagcc	cataatctct	960
acccttacta	ctacaggccc	tacctcaat	atcataggcc	cagtccagac	taccacaagc	1020
cccacccaca	ctatgccaa	cccttcttcc	cacagcaata	gtccccata	cgtagatttt	1080
tgctcttctg	tatgtgacaa	cattttttgt	cattatgtta	ttggaatttt	ctttcataca	1140
ttatattcct	ctaaaactct	caaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1200
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaa		1248

&lt;210&gt; 102

&lt;211&gt; 1841

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 102

ggcacgagct	ggtgcaggag	ctggagcagt	accagttggt	gccgaagaga	ttggactggg	60
agggcaacga	gcacaacagg	agctacgagg	agttggctct	gtccaataag	catgtggctc	120
ctgatcatct	tttgcaaatc	tgccagcgca	tcggctctat	gttgataaa	gaaattccac	180
ccagtatttc	aagagtcact	tctttacttg	gtgcaggaa	gcagtctttg	ctacgtacag	240
caaaaggtag	cttaatttga	agaagtgttc	tgctttgtat	taatagccta	ataattatta	300
taattgtgaa	gtaatgtaag	catgtggaag	aaaggtgata	atgtaactat	aaatcatgct	360
gtggactata	aaattctctg	catttagcat	agaacattga	attttattac	ataagtgttt	420

```

aatggggtcct tctgcattaa aaaagttgat aaagatatag atgttaagct aaaaagcttg 480
aattttattct gagatgttgt ttgagggttt atcaatgtga gttcttaatt ttccattttg 540
tagctcactc ataacttcat tgaatgtaat tgtatgtgtc taataaactc tttaaagtct 600
agagggtattt ctttgatat tttagcttctg aaagcccttt gttataaagg gaagtgtattt 660
tctaaatttg gtagcagctt gaaaacaagc accacaaatg taaaattgca tggttctttt 720
tttactaatt taattttttt gccaaagttat gtgctagcaa aatgtttgaa attgctattc 780
tttttctgtt agcttggtat atatgactta gactaagtat atgtatgtga tataatatat 840
gtgtacatgt ggctaaactg ctaatgggca aagattgttg gttcttctga tgctgagtat 900
acatcaagtt catcccatga ttattttttt ttacgagcta ttgctgatgt gatgctactc 960
ttagctgcac caccagtata aagcttaact gagaaagaac accctccct cttagtcccg 1020
tgccagactg gtttgttagc tttagtgtt tcttagtgct ctagccatgt caccctgttg 1080
cactaatcag gacactgttt atattcttcc tcttctgtgt cctattcac tttgtaatta 1140
ttctatttta gctggttata tggtagttaa ttggtagcct gttttctttt gttttcttta 1200
cttggtgcaa gctccatgtc gttgttttgt gatgagcagt gatttagaac tggccagacc 1260
ttagaagagt cttcagcaag ctgccctagg gtaggggcat attctctgtt tgttttgtac 1320
tggtagccca ggaggatgcc ttgtaactgc tctttgggtc tgatagtggg aatgaatcca 1380
ttcatgtaaa atgcaatatg tctgttagtt tggggcaccc gaggctcagt agagggttag 1440
tatgatgcct tccctttctt ctttgagtt tgaagaaact aatttctggc acagtgtgac 1500
tgcttcaaac ctatgttctg ctgctatctc ctatagagga gggcagactt tttttcttta 1560
gtgcatttgc tggtaggtca aggagggatg gactttgaat tagtgttcag agtcacagat 1620
cctacgtatg tgctcttcag tagaggattt tctgtgatcc tacaatgaag ggaaagctat 1680
ataattttatt ggtacttcct aaagatagag actaaagtca tggtagtatt ggccaaatta 1740
attttaaaaac tgaacaagtc atgaattttc atataaatca aagaactttt tccttataaa 1800
agcttccatt tttgaaaagt tgaaaaaaa aaaaaaaaaa a 1841

```

<210> 103  
 <211> 685  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (678)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (679)  
 <223> n equals a,t,g, or c

```

<400> 103
actacggctg cgagaagacg acagaagggg ggcggcgacg gaggaggagg atggaggcgg 60
tgggtgttcgt cttctctctc ctgcattgtt gcgcgctcat ctctctctcg gtctacttca 120
taattacatt gtctgattta gaatgtgatt acattaatgc tagatcatgt tgctcaaaat 180
taaacaagtg ggtaattcca gaattgattg gccataccat tgtcactgta ttactgctca 240
tgtcattgca ctgggttcac ttctctctca acttacctgt tgccacttgg aatatatatc 300
gatacattat ggtgccgagt ggtaacatgg gagtgtttga tccaacagaa atacacaatc 360
gagggcagct gaagtcacac atgaaagaag ccatgatcaa gcttggtttc cacttgctct 420
gcttcttcat gtatctttat agtatgatct tagctttgat aaatgactga agctggagaa 480
gccgtggttg aagtcagcct acactacagt gcacagttga ggagccagag acttcttaaa 540
tcatccttag aaccgtgacc atagcagtat atattttct cttggaacaa aaaactattt 600
ttgctgtatt tttacatatt aaagtattta aaaaacatga aaaaaaaaa aaaaaaaa 660
aaaaaaaaa aaaaaaanna aaaaaa 685

```

<210> 104  
 <211> 1168  
 <212> DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 104

ggcacgagcc	cccgccectc	ccctccctcc	cctcccttcc	ccggcccccg	ctctggcccc	60
ggcccattcg	ctgttggttc	ttctgctagg	gaggatgtcg	ggttcgtcgc	tgcccagcgc	120
cctggccctc	tcgctgttgc	tggctctctg	ctccctcctc	ccagggccag	gcgccgctca	180
gaacgtgaga	gtacaatctg	gacaggatca	gaagtagaga	atgaagtgtg	aagagaaagg	240
gaaagacaga	agaaaggctg	cagtagtaca	aggagaaaag	caggatgcaa	gaatgaggaa	300
tgaatctttg	tttgaggagc	atccggaaaa	atataagctg	tcagaaagag	taaataagacc	360
agggacctct	aaagtaaatt	cacacatcaa	agttaaaata	atggttgaga	atcacctcct	420
gtgaaaatt	ggcttagctt	tcagtatgcc	tctttaaaaca	aaacattatc	atttttatata	480
aattttttaa	atgttggtca	taatattggag	tttaattttta	agcaccttag	ataatgcatt	540
tgttagcttg	tttgtgaaat	ttttatggat	tttttttttc	aaaattccta	attttagttg	600
gtaaggatta	acttcgggaa	gacaggaaac	ccctccagta	aaattaattg	gttataaatg	660
gttacatatt	ttaggcttat	atacataaca	aaacttctaa	ctgaatttaa	aagtgatctt	720
gtgtaaaaac	attattctag	tgatattgat	gtctaaattt	aaaagagtga	atacacaagt	780
aaaatatatc	tgctttttta	ggatctattc	agtttaggaa	ggagagtcaa	aatttgtatt	840
taattttcaat	ttatttatatt	gtctgaaact	gaaagtaggc	ctaacttttg	tttgcttttg	900
tgtatgtaca	aaggcaaaac	tttatcaatc	aactcttatt	aatttgagat	tattttgacc	960
tgattgtcca	gaacttttgt	ctatctgtat	agaaatgggg	ttttctaaat	atttaaagaa	1020
tttctacgt	atgtaaattt	acttgcaaac	agtaagttag	gggagacatt	taataattcct	1080
cagataacctg	ctgttttctt	gttgtgtttc	ttgtttttca	aataaataaa	ctgagtgtta	1140
tctgttcatt	aaaatagaat	gtagtcag				1168

&lt;210&gt; 105

&lt;211&gt; 1175

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (24)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 105

gcatggaaat	atatectata	catnaagtgc	tgtcctgagc	tgtacatwta	gccatattta	60
gratttttatt	argcaggatt	caaggcctcc	taaaaatgga	tattcagata	tttaaaaaata	120
cactattaag	gaatgcaatg	taaaatcaca	ctaacatttg	catagcactt	tatagttttac	180
aaaatgcctt	cacatatacg	agctcattta	tttctccttc	tcttctttct	attcatctat	240
caaggaatat	cttctataag	ccaggcatca	ggcttgacac	tgaagacaca	aaatgaaaaa	300
gatatacaag	tctctattct	taaagagttc	gtagtttaat	gaagtgacct	aaaaagttaa	360
cagttgcaat	ataacatgca	aaggagtgtg	atgaaataca	cctagagtag	tatgaaaatg	420
gcatagaggg	accaagcctg	taatgaagtc	agagaaagta	aaagaggtga	agcckgagct	480
aagtttttaa	ggatgaacaa	aaattgacca	gacaaaggag	ttggagaagt	gggatgggtac	540
aggataaaa	ccttttcta	ttgggggaagt	gtgagtcgtc	gatattagat	gatattatta	600
atagattgtg	cacagagtgc	acatggtaga	gaggtgagag	atgaggtagc	actgttaggg	660
cagatagggt	aattctcaca	gtaaccctgt	gwtgtacmtt	atcatcctcc	tctccattat	720
gtaggggaga	acaccaaagc	ccaaagagat	taagtaactt	tccagaattt	tgatttcaaa	780
tcctaagcgt	ttggtagctg	tgtttccctg	catatagtct	gcctcaaaag	aaaattttaa	840
gttgattttt	gtattaaatt	taaacaaata	ragggccggg	gcagtggctc	acgcctgtag	900
tcccagcact	ttggtaggcc	aagggtggcg	gatcacctga	ggtcaggatt	tcgagaccag	960
cctggccaat	atgggtgaaac	cccattctcta	ctaacatata	aaaattagct	gggcatgggtg	1020
gcacatgcct	gtaatcctgg	ctacttggga	agctgaggca	ggagaatctc	ttgaaccagg	1080
gaggcggagg	ttgcagttag	ccgagattgt	gccactgcac	tacaacctgg	acaacagagc	1140
aagattccgt	ctcaaaaaaa	aaaaaaaaaa	ctcga			1175

&lt;210&gt; 106

<211> 1021  
 <212> DNA  
 <213> Homo sapiens

<400> 106

ggcagcagaa	taaaatagat	ttaggggtatt	cgacacattt	aatacataat	gaatagtgaa	60
tttcttgaat	aatataatgc	ctaaatgtga	ctgatatgca	aaatatttaa	atcccttaca	120
ggatgtgcat	ttaccaatca	gaacagatgt	tggctttact	gcttgtatta	gttttctgta	180
ttagtttgct	tgtattagtt	tgctggggca	gccataacaa	agtaccacag	aaattttatat	240
tctcacaatt	ctggggggcta	gaagactgag	ataaagatgt	ctgcagggtt	ggtttcttct	300
aaggactctc	ttgttggctt	gtagatgggt	gttgtctccc	tgtgtcttca	cctgcacttc	360
cctctgtacc	tgcccgtatc	ttaatctctt	cttataagga	cagagtcata	ttggattagg	420
gcctcaccct	aatgacctcc	ttttaactta	attacctctt	tagagactct	gtcttcaaatt	480
acaggcacat	tctgaggtac	tttaagggtta	gaacttcaac	atatgcattt	ggaactgggg	540
cataactgag	tccatgacac	tattggagag	ggttctggag	ccataacaga	tggaatgaat	600
gggccttggt	tggattctaa	tttaataaaa	ccaactatag	aaaaacatgt	tttaggcaat	660
cagatttttt	gttatgggct	aaatatattag	aagatgttaa	gaaatgttta	ttaatctcat	720
taggtaaaat	aatgttgggt	tggttatgta	gagtatgact	tacgtgcaca	gaacagaatg	780
ataaggtgtc	caggatttgc	tttaaaatat	tctagatttt	tttaaaagggt	gagaggaaat	840
gacacatgta	tatctaaatt	ttgatgggtg	ttgaagctgg	ctggtaggta	cctagtattt	900
cattatactt	tttcctttgt	gtgtttggaa	atttttataa	taagcaggaa	aaataaaagt	960
aagaatatat	gtattataga	taatttgaaa	tatgtagaat	catgaaaaaa	aaaaaaaaaa	1020
a						1021

<210> 107  
 <211> 830  
 <212> DNA  
 <213> Homo sapiens

<400> 107

ggcagcagta	aggactgtgt	tctttatgca	tttcttgatc	caggcatggc	agttcctctt	60
ttcctgtaca	tattcacact	cctgccactt	ctaccctttc	tcttatccct	ctgcttttca	120
cctctgactg	taaaaagaag	tagcagttcc	gaaagcaaga	gttccctatg	aacacggaag	180
aagacattgg	caacttttga	gtacaacaac	tatatttaat	agagtaattt	aagaacatca	240
gccagtgaat	tttatacaag	atagtgaag	agaaaaggaa	gattaattag	gggtagttta	300
ggatgccatt	aaatagccta	gaattagggg	agtagtcgtt	gaatagaaaag	gaggccacaa	360
atttgagggg	tataagctaa	gaattggtaa	gcgaagaaga	aggaaaagggt	ttgggcagta	420
aggataatga	ggaacaaaat	agagaactca	gaagcaatat	ctgactgtta	tcattggaag	480
aatttttttt	cttgcttgag	gctggatatt	gaagtggatc	aggataactg	agtgactatc	540
tgatgggctt	ttggaactag	ctctcaagag	gtgaaaatta	gctttttttt	ctttttcttt	600
cttttttttt	ttttttgagg	caaggtctca	ctgttggtga	ggctgaacct	cctgggctca	660
agcagttgtc	ccattgcagc	ctctcagat	actctgtaag	ccaaggcagg	gggaatattt	720
tgtgtcagc	agtttgaggc	tgtgggtgagc	taagatcaca	ctgctgtgct	catttcagcc	780
tgggcaacac	agtgaaaccc	cgtctccatc	tgtttaaaaa	aaaaaaaaaa		830

<210> 108  
 <211> 1301  
 <212> DNA  
 <213> Homo sapiens

<400> 108

aagaggggtg	ctgccgagtt	cttgtttgtc	ctgtgctctg	agagtgggtga	gttatggaga	60
cccaggtccc	agcccccccc	acctcagcca	ggcttgccaca	ctgaacctctg	ccttgccccc	120
cagtgcctcg	attcatcaag	tacacaggct	atgggaatgc	tgctggcctt	ctggctgccca	180
ggggcctcat	cagcaggagg	cggcccagag	gccagtactc	agaggatgag	gacacagaca	240
cagatgagta	caaggaagcc	aaagccagca	taaacctgtg	gaccgggagg	gtggaggaga	300
agccgcctaa	ccctatggag	ggcatgacag	aggagcagaa	ggagcacgag	gccatgaagc	360

tggtgaccat	gtttgacaag	ctctccagga	acagagtcac	ccagccaatg	gggatgagtc	420
cccggggtca	tcttacgtcc	ctgcaggatg	ccatgtgcga	gactatggag	cagcagctct	480
cctcggaccc	tgactcggac	cctgactgag	gatggcagct	cttctgctcc	cccatcagga	540
ctgggtgctgc	ttccagagac	ttccttgggg	ttgcaacctg	gggaagccac	atccccactgg	600
atccacaccc	gccccactt	ctccatctta	gaaacccctt	ctcttgactc	ccgttctgtt	660
catgatttgc	ctctggtcca	gtttctcatc	tctggactgc	aacgggtctt	ttgtgctaga	720
actcaggctc	agcctcgaat	tccacagacg	aagtactttc	ttttgtctgc	gccaaagagga	780
atgtgttcag	aagctgctgc	ctgagggcag	ggcctacctg	ggcacacaga	agagcatatg	840
ggagggcagg	ggtttgggtg	tgggtgcaca	caaagcaagc	accatctggg	attggcacac	900
tggcagagcc	agtgtgttgg	ggtatgtgct	gcacttccca	gggagaaaac	ctgtcagaac	960
tttccatacg	agtatatcag	aacacaccct	tccaaggtat	gtatgctctg	ttgttctctg	1020
cctgtcttca	ctgagcgcag	ggctggaggc	ctcttagaca	ttctccttgg	tcctcgttca	1080
gctgcccact	gtagtatcca	cagtgtcccga	gttctcgctg	gttttggcaa	ttaaaccctcc	1140
ttcctacttg	tttagactac	acttacaaca	aggaaaatgc	ccctcggttg	accatagatt	1200
gagatttata	ccacatacca	cacatagcca	cagaaacatc	atcttgaat	aaagaagagt	1260
tttgacaaa	aaaaaaaaa	aaaaaaaaa	aaaaaactcg	a		1301

&lt;210&gt; 109

&lt;211&gt; 1932

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 109

aatttttttt	tttttttttt	tttttttgaa	aaaaaaaaatg	ggtagtgtat	atttttgcagg	60
tttaagacaa	ctcaggacaa	taaaaacaat	ggactttaca	tgtgtatata	tatagctctc	120
ttaggcacca	taatcagtat	gagccaacaa	tattttaaact	tgattcaggc	cacattcaga	180
catttgctct	tatatacaaa	tattttaaatt	aaatacaatc	tgaaatgtgt	tctgttacat	240
acaaaaaagg	aaaaactata	caacgcagag	cagtgtgtgt	gtttttaata	attacattta	300
catgtaagct	aaatggaacc	agcaatgggt	ctcaagtttt	tatcatccct	tccagaaaat	360
ctttttctac	catctcttct	attttttgcc	tggctttgct	ggaacatggg	ttgtgggtct	420
ccagtttcat	gtccttatta	gggaaggcat	ttgagtagag	gataggactc	cctgagtgtc	480
ctccacatcg	gcttgtgact	ttgctgttga	agacttgact	gagcacattg	aagaacggca	540
ggagctgctc	catactgcgc	acggtgcaga	tggtagcag	caagtgcctt	ggctcccaac	600
ccaatgttct	ccttgagttg	tcttctctct	gatttttctg	cagaaaacaa	aaagtgaact	660
ggtattaata	caacagacaa	tgtggtatgt	tagaaaaatt	aaaaatatat	aaactttggc	720
aattgggtcaa	gaaatgaata	caaatgacat	taagtttcta	actcctgacc	tgatcaaaac	780
ccttgggtgct	tctgagacct	tttactgcca	tttattagtt	ttacatggag	cagtctaaca	840
ttgtagtaat	agttcccaac	tagaatgcgc	agataagctt	agttaacaga	aatagctttg	900
aacaggaata	gagtcaaaac	taaaagtttt	atggttgtgt	ttgtattttac	tcaaaaagct	960
cccaggtttc	tgaacctca	ctactgtaac	caaggactag	gtcacaaaat	tactacagaa	1020
aaaaggaaca	aagtgtctta	tacatttcat	aatatatccc	cttttattat	aattagttaa	1080
ttccctttta	tctaaatggc	ctaaatttgc	catgatggta	gcagtgtcca	aagtgaataa	1140
ttactgtcag	tactgcatca	cagagaaaagg	aagggatccc	tcaggagaca	ctgctgtctc	1200
cttctggggt	gtgctaaaca	acatagggag	gaaagctgga	cctggagtca	aaggaattga	1260
gttagtgtgc	tggctctgcc	atacttacgg	cacccttggg	caggatatac	aaaggttccct	1320
cacttataaa	atgggacagt	ctaaaactac	cttttagtag	agaagtcaaa	tgagaaggta	1380
tgtgaaaact	ctgtcaacta	aatataaaga	ctaataat	gggtattaa	aggctagttt	1440
gagaagccac	ctgaattaca	caaacacagc	tacagacatc	attctgtcta	gagaaagata	1500
agagagaaca	ggttgggtga	acttgggcag	acaacacagat	acaattccac	actaaagaat	1560
gaaaataagc	aatgaactag	acagaaggaa	gaaatcatga	agacttagga	agcagaatta	1620
caatctgtca	tattaacaaa	tggagtgtgc	cttctaagat	cagatgttgc	tcagaaactt	1680
tcattgttta	cctaataatt	taatatcact	agtttcttag	tgggtcaagc	agatgcaaaa	1740
tccagcttat	tttcttctat	gtgctctcaa	gcttattgct	tattttaaag	taaaatcctg	1800
aaaaaggaaa	atattaggtt	ggtgcaaacg	taattgcggt	ttttgcattg	ttgaaatttg	1860
ccgttttata	ttggagtaca	ttcttaataa	aatgtggtta	tggtatacaa	aaaaaaaaaa	1920
aaaaaactcg	ag					1932

<210> 110  
<211> 1534  
<212> DNA  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (1212)  
<223> n equals a,t,g, or c

<400> 110  
tcgaagatag gttcgagacg gtggatgttg cagctgatca tgcagttggg ttcggtgctg 60  
ctcacacgct gccccttttg gggctgcttc agccagctca tgctgtacgc tgagagggct 120  
gaggcacgcc ggaagcccga catcccagtg ccttacctgt atttcgacat gggggcagcc 180  
gtgctgtgcg ctagtttcat gtcctttggc gtgaagcggc gctggttcgc gctggggggc 240  
gcaactccaat tggccattag cacctacgcc gcctacatcg ggggctacgt ccactacggg 300  
gactggctga aggtccgtat gtactcgcgc acagttgcca tcatcggcgg ctttcttgtg 360  
ttggccagcg gtgctgggga gctgtaccgc cggaaacctc gcagccgctc cctgcagtc 420  
accggccagg tgttcctggg tatctacctc atctgtgtgg cctactcact gcagcacagc 480  
aaggaggacc ggctggcgta tctgaacctc ctcccaggag gggagctgat gatccagctg 540  
ttcttcgtgc tgtatggcat cctggccccct ggcttttctg tcaggctact acgtgacctt 600  
cgctgccag atcctggctg tactgtgccc cctgtcatg ctgctcattg atggcaatgt 660  
tgcttactgg cacaacacgc ggcgtgttga gttctggaac cagatgaagc tccttgga 720  
gagtgtgggc atcttcggaa ctgctgtcat cctggccact gatggctgag ttttatggca 780  
agaggctgag atgggcacag ggagccactg agggtcaccc tgccttctc cttgctggcc 840  
cagctgctgt ttatttatgc tttttggtct gtttgtttga tcttttgcct ttttaaaatt 900  
gttttttgca gtttaagaggc agctcatttg tccaaatttc tgggcttcag cgcttgggag 960  
ggcaggaacc ctggcactaa tgctgtacaa ggttttttct ctgttaggaa gaacttgagg 1020  
ccagctcccc actgagtctt ctgtccctga agaaaggagc tattgggcag ggcttgggat 1080  
ccggctactg agagtgggag agtgggagac agaggaagga agatggagat tggaagtgag 1140  
caaatgtgaa aaattcctct ttgaacctgg cagatgcagc taaactctgc agtagtgttt 1200  
ggagactgtg anaggagtg tgtgtgttga cacatgtgga tcaggcccag gaagggcaca 1260  
ggggctgagc actacagaag tcacatgggt tctcagggtg tgccaggggc agaaacagta 1320  
ccggctctct gtcactcacc ttgagagtag agcagacctt gttctgctct gggctgtgaa 1380  
gggggtggagc aggcagtggc cagctttgccc cttcctgctg tctctgttct tagctccatg 1440  
gttggcctgg tgggggtgga gttccctccc aaacaccaga ccacacagtc ctccaaaaat 1500  
aaacatttta tatagacaaa aaaaaaaaaa aaaa 1534

<210> 111  
<211> 2871  
<212> DNA  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (1234)  
<223> n equals a,t,g, or c

<220>  
<221> SITE  
<222> (1259)  
<223> n equals a,t,g, or c

<220>  
<221> SITE  
<222> (1283)  
<223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (1284)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (1287)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (1378)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (1912)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (1913)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (1935)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (1947)  
 <223> n equals a,t,g, or c

<400> 111

gcggcgaggc	ggcgggtggtg	gctgagtcgc	tggtggcaga	ggcgaaggcg	acagctctag	60
gggttggcac	cggccccgag	aggaggatgc	gggtccggat	agggctgacg	ctgctgctgt	120
gtgcggtgct	gctgagcttg	gcctcggcgt	cctcggatga	agaaggcagc	caggatgaat	180
ccttaggatt	ccaagactac	tttgacatca	gatgagtcag	taaaggacca	tactactgca	240
ggcagagtag	ttgctggtca	aatattttct	gattcagaag	aatctgaatt	agaatcctct	300
attcaagaag	aggaagacag	cctcaagagc	caagaggggg	aaagtgtcac	agaagatata	360
agctttctag	agtctccaaa	tccagaaaac	aaggactatg	aagagccaaa	gaaagtacgg	420
aaaccaggta	gtctggacat	tttccttgct	ttttgattta	tttaggggac	aactgaaaat	480
tttaagctaa	tgaataaaga	agctgaagaa	gactggcttc	actgattatt	acccacaaat	540
aatatatgga	gtgtagtttg	gagagaattt	ctagatttta	atataccaaa	gttattcact	600
taacagattt	gacccaagtt	acataagctg	aattccatag	atgagattgc	atttatcttt	660
ttattaatat	cctgacgttt	ctttttggaa	tcaagtatgt	aaataccttt	ctcactgaga	720
aatttaaaaa	aataatatgt	cagggtagca	tacaaaacag	aaactcagga	agaactgtct	780
aacatgccac	tatggttatt	ttcaggattg	tggtaggatt	attactacaa	taacctgtgc	840
gatcactcta	ggtaatttga	tttaggtggt	tatccctcca	gcttgacgct	tgtgtgtact	900
tccatcctat	tctaactttc	acaatcttga	aacctctact	tttcaattta	aaacttgacc	960
tttttcacaa	tagtcttcta	aatattaatc	cttttaaaaa	ttattaatgt	atgcattcta	1020
ccatgtcttg	ttgcccttct	gtcatttatt	atttcaatga	tcactgcctt	atctgtttat	1080
tattccatta	ctaaccactc	tcaggaactg	atccaaattg	gaacttttta	aaaaaataga	1140
attttttttt	actacaaata	taccattctc	atgaatagaa	aactattttt	aaaaaattaa	1200
aagtacctga	atcccatcaa	ctagaaataa	cagntgctca	catcatggag	aatattctnt	1260
caggattttt	ttaggtgtat	cannagnttt	aagaaaagta	aattgggtca	aaatgtgtat	1320
tctatgttgc	agcctgcatt	ttgcacttaa	cagttttacca	tgaatgtttt	tccatgtnat	1380

ttagcttatt	ttcaatttga	tagttaatgg	ctataaaaaa	ttttatttgt	aaatatatgg	1440
taccataaac	caaaacgttt	atgttttgct	gggagatcat	tttagatgta	tctttgtgcc	1500
tatctgtgaa	aatttcctta	gaattcttaa	agtaaatttg	ctgaattgga	tgtaaaagtt	1560
taaaagacag	tagatacaca	ttgttaaatt	gttcttcaga	aaagtgattc	tgtttacttc	1620
ctctgaaaat	acctctttcc	ctgtacctca	cccaagtgtg	tattaccatt	ttaaaatttt	1680
tactaaatat	aattttaattc	attcttcatt	ttttaattag	ggattttttt	tttttttttt	1740
tagtttttat	atttttagag	acatgggtctc	actctgtcac	tcaggctgga	gtgcagtggc	1800
acgatcatag	ttcactgcag	ccttgaactc	aggatgatctc	cctctttggc	ctcccaaagt	1860
gctgggatta	caggcatgag	acactacatc	aagcctggga	aattttttaa	cnngcataaa	1920
aagtagaggg	taaantgcca	ccccttncca	accctcactc	agcttcaata	aacatcaaca	1980
ttctgctggg	cttttcattc	ttactacct	acacacattt	gccctttttt	tctttcctag	2040
aatattttta	ggcaaattcc	gcatgtgctt	ttacatctct	gtgtctctct	ctagtaagaa	2100
tgtctttttc	agcaaaatca	cagaatttagc	cattactcct	ttattacct	ataccagac	2160
tgtttaattt	ttctagtgtc	ttcaagtgtc	tgtttacaat	ttatttgttt	gaattcattt	2220
ctaagctact	caggaggctg	agtcaggaga	attgcttgaa	ccctagaggc	agagggtgca	2280
gtgagctgag	atcgtgccac	tgtactccag	cctgggtgac	aaagtgagac	tccatgtcat	2340
aacaaaacaa	aacaaaacaa	aacatgacga	aacaaaacaa	acaattgaat	aatatccctc	2400
agtattattt	tagagaagag	atgactcaat	ataaatgtaa	tatgtgatcc	atttggtcac	2460
agaatgaatt	tcagtcattg	cttaggtaag	cctgggtaaaa	aaaaaaaaaa	cgggattcca	2520
tgattagtgt	gttgaaaaag	gggtttgccc	aaagtgtaac	agacagagcc	gggtgcagtg	2580
gctcacgcct	gtaatcccag	aactttggga	ggccaaggca	ggcggatcac	gagggtcaaga	2640
gattgagaca	atcctggcca	acatggtgaa	gccctgtatc	aactaaaaat	acaaaaacaa	2700
ttagctgggc	gtggtgacat	gtgcctgtag	tcccagctac	tcaggaggct	gaggcaggag	2760
aattgcttga	acctgggagg	cggaggctgc	agtgagctga	gattgctcca	ctgcactcca	2820
gcctggccac	agagtggagt	tccatctcaa	aaaaaaaaaa	aaaaaaaaaa	a	2871

&lt;210&gt; 112

&lt;211&gt; 1037

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (936)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (946)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (951)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 112

ccggtccgga	attcccgggt	cgaccacgc	gtccggggcg	gcgcaggacg	tgcactatgg	60
ctcggggctc	gctgcgccc	ttgctgcggc	tcctctgtgt	ggggctctgg	ctggcggttc	120
tgcgctccgt	ggccggggag	caagcgccag	gcaccgcccc	ctgctcccgc	ggcagctcct	180
ggagcgcgga	cctggacaag	tgcattggact	gcagcacctc	ctgccccctt	ccggctgtct	240
tggcccatcc	ttggggggcg	tctgagcctg	accttcgtgc	tggggctgct	ttctggcttt	300
ttggtctgga	gacgatgccg	caggagagag	aagttcacca	cccccataga	ggagaccggc	360
ggagagggct	gcccagctgt	ggcgctgatc	cagtgcacat	gtgccccctg	ccagccgggg	420
ctcgcccact	catcattcat	tcatccattc	tagagccagt	ctctgcctcc	cagacgcggc	480
gggagccaag	ctcctccaac	cacaaggggg	gtggggggcg	gtgaatcacc	tctgaggcct	540
gggcccagg	ttcaggggaa	ccttccaagg	tgtctggttg	ccctgcctct	ggctccagaa	600
cagaaaggga	gcctcacgct	ggctcacaca	aaacagctga	cactgactaa	ggaactgcag	660



catttgcaca	ggggaagggg	gtgccctcct	tcctagaggg	cctggggggc	aggctgactt	720
ggggggcaga	cttgacacta	ggccccactc	actcagatgt	cctgaaatc	caccacgggg	780
gtcaccctgg	gggggttagg	acctattttt	aacactaggg	ggctggccca	ctaggagggc	840
tggccctaag	atacagaccc	ccccaaactc	ccaaagcggg	ggggggatat	ttattttggg	900
gagagttttg	agggggaggg	gatttttttt	ttaaangatt	ttttantttt	naaaaaaaaa	960
aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	1020
aaaaaaaaaa	aaaaaaa					1037

<210> 113  
 <211> 2214  
 <212> DNA  
 <213> Homo sapiens  
  
 <220>  
 <221> SITE  
 <222> (289)  
 <223> n equals a,t,g, or c

<400> 113

cggacgcgtg	ggtcgaccca	cgcgtccggg	aaaaarggaa	aaaratgccgt	gtaaaatctc	60
gttctgtgtc	tgaattgccg	taggctcaga	tcttcatttg	aggttctgtg	tctgaattgc	120
cgttaggtca	gatcttcatt	tgaggttatg	ttctataagt	taacgttgat	cttgtgtgag	180
ctttcggtag	ctggagtaac	acaggcggcc	tcacagcgac	ctctccagcg	ccttccaagg	240
cacatctgca	gccagcgtam	tcctcctggg	agatgcctcc	tcaaggccnt	gctccagacc	300
acgtgggrar	ggcctgacaa	gccaattccc	aggtctgtcc	cacccttgra	gagtgacctt	360
aaacgctaga	cagatgggga	atgggaaaga	aaagaaagct	gcagacctca	agttaaaatt	420
ccctcaaaaa	cgttttttatt	tatctgcttt	ttctgaaagg	ataaaggctt	tttgaaaatt	480
attttctaac	aaataacatg	aacacttcta	gaaaccctag	aaaaacacaa	agtattcaaa	540
atagaaagaa	aaattaccca	ttactcttta	agccagcatt	atccattgcg	gtgcttttgg	600
agttgggtga	ggcctgagcc	tctgccaaat	caaggagccc	gggtggtggc	gtggcattcc	660
tgcagggttg	tttttttttc	tttgagatgg	agtctcactc	ttgtcacccc	agctggaatg	720
tggtggtgta	aacagctcac	tgcagccttg	accctgaggg	tcaagcgatc	cttctgcctt	780
ggcctcctga	gtagctggga	tcccaggcga	gagtcaccac	accctgtcca	tgttcctgca	840
ggcttctgata	tgcaggagac	ctgtgtcttc	cctgccacat	tttcttcttc	tttcttgaga	900
cagacccttg	ctccatcacc	caggecagag	tgtggtstgt	cgaacacggc	tcactgcagc	960
ctcgaccctc	aggctcaagc	gatcctcagc	cctcggaacc	ccaaagtgc	gggatcacag	1020
gcgagagtca	ccatgctggc	ctgaatcttc	agggattttr	cggttgargt	gycacttact	1080
tarccatssc	tgtttcaaga	gtgtaggtgg	tcaccctgtc	tctgccgctg	acctggcctg	1140
gaccctcggc	tgtgagaggg	aggggtgggc	tgggctggag	gaacctraag	ccctcgtgat	1200
gtcacaagcc	catctggctg	ggcatccctt	gctgtgtcct	gagctgcaca	tgccccaggt	1260
ggccccaca	gcagaggcga	gccactgrag	ggtgragggc	ttccacggac	ggtcttcagg	1320
ggragaagaa	ggggccaggc	ccccaggaga	ctcaggagac	cagagcctgg	ggtcaggggc	1380
tmagcagggg	ctyarccagg	gctggatgtc	cggagccagc	cccgmagccc	tgkgktcttt	1440
gttcttcgca	ctcccaccgt	ccgtgtgaac	agctccagcc	ccacctgcgc	ctccctgtgc	1500
tgggctccat	cagggagccc	agaagacgtg	tgtgcttctg	aaattgggtc	cctacatgcc	1560
tttgtccag	tgcaccttgc	tccttccatt	tactatcgag	atttaaatgc	ctgttttctc	1620
cccagagggt	gacggatata	ttcagacggt	acgacacgga	tcaggacggc	tggattcagg	1680
tgctgtacga	acagtacctg	tccatgggtc	tcagtatcgt	atgaccctgg	cctctcgtga	1740
agagcagcac	acatggaaga	gagccaaaat	gtcacagttc	ctatctgtga	gggaatggag	1800
cacaggtgca	gttagatgct	gttcttccct	tagattttgt	cacgtgggga	cccagctgta	1860
catatgtgga	taagctgatt	aatgggtttg	caactgtaat	agtagctgta	tcgttctaat	1920
gcagacattg	gatttggtga	ctgtctcatt	gtgccatgag	gtaaatgtaa	tgtttcaggc	1980
attctgcttg	caaaaaaatc	tatcatgtgc	ttttctagat	gtctctgggt	ctatagtgca	2040
aatgctttta	ttagccaata	ggaattttta	aataacatgg	aacttacaca	aaaggctttt	2100
catgtgcctt	acttttttaa	aaaggagttt	attgtattca	ttggaatatg	tgacgtaagc	2160
aataaagggg	atgttagacg	tgtaaaaaaa	aaaaaaaaaa	aaaaaaaaaa	aaag	2214

<210> 114  
 <211> 3300  
 <212> DNA  
 <213> Homo sapiens

<400> 114

tcgacccacg	cgtccgggtga	gaggcaagga	cttttcatct	ttagggcatt	ctgaccacgt	60
cctgcttcag	agagattggt	cccggcgtct	cagtgcctatg	gggagcaggt	tcctcctggt	120
cctgctctca	ggtctcactg	tcttactggc	tctgccagga	tcagaagcca	agaattctgg	180
agcttcctgt	cctccatgcc	ctaaatatgc	cagctgccac	aacagcacc	actgtacttg	240
tgaagatggc	tttcgggcca	ggtctggcag	gacatacttt	catgattcct	ctgagaagtg	300
tgaagatatt	aatgaatgtg	aaaccgggct	ggcaaagtgc	aagtataaag	catattgtag	360
gaataaagtt	ggagggttaca	tctgtagctg	tttggtaaaa	tatactttat	tcaactttct	420
ggctgggtatt	atagattatg	atcatccgga	ttgttatgag	aacaatagtc	aagggacgac	480
acagtcacac	atggatattt	gggtgagtgg	ggtgaagcct	ggatttggga	aacagctggt	540
acgtataact	atgccatttt	cctacccaaa	cattaacatg	tcttcctgtg	atttttaggg	600
tagggtagtt	ctatccaggg	gtaattttgt	cctctgtccc	aaggtcatct	gtcaatgact	660
ggggacactt	ttggttgtca	taccttgggg	gtgatgtgtg	tgactggcat	ctggtggatg	720
gagaccaggg	atacagctca	acatccctaca	gtgccagga	cagcctccca	caatcaagaa	780
gtgcctagt	ccatatgtcc	atagagatag	agaaatacaa	gtgtaggggg	aaagtgcctc	840
agctggcatt	caaagaccta	catcaagcac	ctagattctc	aatgccacac	gcaccttgta	900
gcaccttaat	aaatatcggt	gccttctggg	ctcccactc	caacacttgt	gcataattccc	960
tatttcttac	atttcagtaa	gagtcaatct	aattagctta	atttttttgg	aaggaaaatc	1020
tgagaagaaa	tggaagcaga	gaggactttg	caagaagggc	tactcaacta	attcaaagcg	1080
tggagttag	catctggaat	gcgagttttg	cttctccagg	aaaggggtcaa	atttctgaat	1140
ttgatatagt	ctatgaaacc	aagaggtgca	atgagacaag	ggagaatgct	tttctggaag	1200
ctggaaataa	caccatggat	atcaactgtg	ctgatgcttt	aaaaggaaac	ctaagagaga	1260
gcactgcagt	tgccctatca	cttatcaatc	tcttggggat	attctgaatg	catecttttt	1320
tagtaaacga	aaagggatgc	aggaagtaaa	actgaactct	tacgttgtga	gcggcaccat	1380
cggtttgaag	gaaaaaattt	ccctctctga	acctgtgttc	ctgacttttc	gccataatca	1440
gcctggtgac	aagagaacaa	aacatatctg	tgtctactgg	gagggatcag	agggaggccg	1500
ctggtccacg	gagggctgct	ctcatgtgca	cagcaacggt	tcttacacca	aatgcaagtg	1560
cttccatctg	tccagctttg	ccgtccctcg	ggctcttgcc	cccaaggagg	accctgtgct	1620
gaccgtgac	acccaggtgg	ggctgaccat	ctctctgctg	tgccctcttc	tggccatcct	1680
caccttcctc	ctgtcccgcc	ccatccagaa	caccagcacc	tcctctccatc	tagagctctc	1740
cctctgcctc	ttcctggccc	acctcctgtt	cctgacgggc	atcaacagaa	ctgagcctga	1800
ggtgctgtgc	tccatcattg	cagggctgct	gcacttcttc	tacctggctt	gcttcacctg	1860
gatgctcctg	gaagggctgc	acctcttctc	caccgtcagg	aacctcaagg	tggccaacta	1920
caccagcacg	ggcagattca	agaagaggtt	catgtaccct	gtaggctacg	ggatcccagc	1980
tgtgattatt	gctgtgtcag	caatagttgg	acccacagaat	tatggaacat	ttactcactg	2040
ttggtcaag	cttgataaag	gattcatctg	gagcttcatg	gggccagtag	cagtcattat	2100
cttgataaac	ctggtgttct	acttccaagt	tctgtggatt	ttgagaagca	aactttcctc	2160
cctcaataaa	gaagtttcca	ccattcagga	caccagagtc	atgacattta	aagccatttc	2220
tcagctattt	atcctgggct	gttcttgggg	ccttggtttt	tttatggttg	aagaagttagg	2280
gaagacgatt	ggatcaatca	ttgcatactc	attcaccatc	atcaacaccc	ttcagggagt	2340
gttgctcttt	gtggtacact	gtctccttaa	tcgccaggtt	cgaatggaat	ataaaaagtg	2400
gtttagtggg	atgcggaaag	gggtagaaac	tgaaagcact	gagatgtctc	gctctactac	2460
ccaaaccaa	acggaagaag	tggggaagtc	ctcagaaatc	tttcataaag	gagggcactgc	2520
atcatcatct	gcagagtcaa	ccaagcaacc	gcagccacag	gttcatctcg	tctctgctgc	2580
ttggctaaag	atgaactgac	ctggcaagtg	ccatggcaat	gaccocggaag	ttacccctcc	2640
tttccgtttg	tctacagcgc	ccctgtgggtc	acacatagat	tggacaaatg	ccactatttc	2700
tagctttcct	gtgaaaagtc	taggctcatt	cacctatttt	ggctttttat	gttcatagaa	2760
agaacaagac	atttgggaga	attcttagat	ccagagtcca	gtagtgtggc	acgtgcaatg	2820
aagtgtcgga	aggatgcatt	ttaaagatgg	cgggcgggag	aagtggattt	tcttcttgca	2880
gctactgcca	ccttgccaga	aactcactaa	ctggcatctg	gattcagctc	atagttccct	2940
ttctggcctc	ctgtgtgtat	tttatgtctc	caaagatctt	acattaacac	tcacatttca	3000
cataattcaa	caattttcat	atggatcagt	attaaagagg	gtgttgcat	ttgcaatata	3060
aaaatgcatt	atcaggtgct	ggagaggatg	tggagaaata	ggaacacttt	tacactgttg	3120
gtgggactgt	aaactagttc	aaccatcggtg	gaagtcagtg	tggcgatttc	tcagggatct	3180

agaactagaa ataccatttg acacagctat cccattactg ggtatataacc ccaaggggcta 3240  
 taaatcatgc tgctataaag acacatgcac acgtaaaaaa aaaaaaaaag ggcggccgct 3300

<210> 115  
 <211> 1286  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (1149)  
 <223> n equals a,t,g, or c

<400> 115  
 gactgcttta cggacattgg atgaagccga agcattttaga atggtgcctg gcacacagtt 60  
 ggtgcgtgat atggttaagc tttgtgtccc caccacatc tcatcttgaa tgtgacggtt 120  
 tccccggctc cctcctgccc ccatgtgaag aagggtcgtt ctcccccttc acctccacc 180  
 accatgattg ccatggatgc tccccactcc aaagcagccc tggacagcat taacgagctg 240  
 cccgagaaca tcctgctgga gctgttcacg cacgtgcccg cccgccagct gctgctgaac 300  
 tgccgcctgg tctgcagcct ctggcgggga cctcatcgac ctcatgacct tctggaaacg 360  
 caatgcctgc gagagggcct catcaccaag gactgggacc agcccgtggc cgactggaaa 420  
 atcttctatt tectacggag cctgcatagg aactcctgcg caaccctgtg gctgaagagg 480  
 atatgtttgc atggcaaatt gatttcaatg gtggggaccg ctggaagggtg gagagcctcc 540  
 ctggagccca cgggacagat tttcctgacc ccaaagtcaa gaagtatttt gtcacatcct 600  
 acgaaatgtg cctcaagtcc cagctggtgg accttgtagc cgaggggctac tgggaggaggc 660  
 tactagacac attccggccg gacatcgtgg ttaaggactg gtttgctgcc agagccgact 720  
 gtggctgcac ctaccaactc aaagtgcagc tggcctcggc tgactacttc gtgttggcct 780  
 ccttcgagcc cccacctgtg accatccaac agtggaacaa tgccacatgg acagagggtct 840  
 cctacacctt ctcagactac ccccggggtg tccgctacat cctcttccag catgggggca 900  
 gggacaccca gtactgggca ggctggtatg ggccccgagt caccaacagc agcattgtcg 960  
 tcagccccaa gatgaccagg aaccaggcct cctccgaggg tcagcctggg cagaagcatg 1020  
 gacaggagga ggctgccccaa tcgccctacc gagctgttgt ccagattttc tgacagctgt 1080  
 ccatcctgtg tctgggtcag ccagagggtc ctccaggcag gagctgagca tgggggtggg 1140  
 aagtgaggnc cctgtaccaaa gcgacttctg ccccggttca accctaccaa gcttggggga 1200  
 acttactgca catagctctg acgttttgtt gtaataaatg ttttcaggcc gggcaaaaaa 1260  
 aaaaaaaaaa aaaaaaaaaa aaaaaa 1286

<210> 116  
 <211> 2189  
 <212> DNA  
 <213> Homo sapiens

<400> 116  
 ggaattcggc acgaggcgcc cgaggatgtg ctgctggccg ctgctcctgc tgtgggggct 60  
 gctccccggg acggcgggcg ggggctcggg ccgaacctat ccgcaccgga cctcctgga 120  
 ctcgaggggc aagtactggc tgggctggag ccagcggggc agccagatcg ccttcgcct 180  
 ccagggtgcg actgcaggct acgtgggctt cggcttctcg cccaccgggg ccatggcgct 240  
 cgccgacatc gtctgtggcg ggggtggcca cgggcgggcc tacctccagg attattttac 300  
 aaatgcaaat agagagttga aaaaagatgc tcagcaagat taccatctag aatatgccat 360  
 ggaaaatagc acacacacaa taattgaatt taccagagag ctgcatacat gtgacataaa 420  
 tgacaagagt ataacggata gactgtgtag agtgatctgg gcctaccacc atgaagatgc 480  
 aggagaagct ggtcccaagt accatgactc caataggggc accaagagtt tgcggttatt 540  
 gaatcctgag aaaactagtg tgctatctac agccttacca tactttgatc tggtaaataca 600  
 ggacgtcccc atcccaaaca aagatacaac atattgggtg caaatgttta agattcctgt 660  
 gttccaagaa aagcatcatg taataaagggt tgagccagtg atacagagag gccatgagag 720  
 tctggtgcac cacatcctgc tctatcagtg cagcaacaac tttaacgaca gcgttcctgg 780  
 aatccgggca cgaattgcta tcaccccaac atgcccgatg cattcctcac ctgtgaaact 840

gtgatttttg	cctgggctat	tggtggagag	ggcttttctt	atccacctca	tggtggatta	900
tcccttggca	ctccattaga	tccgcattat	gtgctcctag	aagtcatta	tgataatccc	960
acttatgagg	aaggcttaat	agataattct	ggactgaggt	tattttacac	aatggatata	1020
aggaaatatg	atgctggggt	gattgaggt	ggcctctggg	tgagcctctt	ccataccatc	1080
cctccagggg	tgcttgagtt	ccagtctgag	ggcactgca	ctttggagtg	cctggaagag	1140
gctctggaag	ccgaaaagcc	aagtgggaatt	catgtgtttg	ctgttcttct	ccatgctcac	1200
ctggctggga	gaggcatcag	gctgcgtcat	tttcgaaaag	ggaaggaaat	gaaattactt	1260
gcctatgatg	atgattttga	cttcaatttc	caggagtttc	agtatctaaa	ggaagaacaa	1320
acaatcttac	caggagataa	cctaattact	gagtgtcgct	acaacacgaa	agatagagct	1380
gagatgactt	ggggaggact	aagcaccagg	agtgaatgt	gtctctcata	ccttctttat	1440
tacccaagaa	ttaatcttac	tcgatgtgca	agtattccag	acattatgga	acaacttcag	1500
ttcattgggg	ttaaggagat	ctacagacca	gtcacgacct	ggcctttcat	tatcaaaagt	1560
cccaagcaat	ataaaaaacct	ttctttcatg	gatgctatga	ataagttaa	atggactaaa	1620
aaggaagggt	ttctcttcaa	caagctgggt	ctcagcctgc	cagtgaatgt	gagatgttcc	1680
aagacagaca	atgctgagtg	gtcgattcca	aggaatgaca	gcattacctc	cagatataga	1740
aagaccctat	aaagcagaac	cctttggtgt	gtggcacgtc	ttcttctctt	tcctgcaca	1800
gagattttct	ccatcaactt	gcttggttgc	cttctgctac	tcagctgcac	gctgagcacc	1860
aagagcttgt	gatcaaaatt	ctgttggtg	tgacaatgtt	ttctatgatc	tgaacctgtc	1920
atgtgaagta	caggttaaag	actgtgtcca	ccttgggcat	gaagagtgtg	gagacttttc	1980
ttccccat	ttccctccct	ctttttcctt	tccatgttac	atgagagaca	tcaatcaggt	2040
tctcttctct	ttcttagaaa	tatctgatgt	tatatataca	tggtcaataa	aataaaactg	2100
gcctgactta	agataaacat	tttaaaaaat	tgggtgtgca	tggtgggaata	aaagaattct	2160
ttctttccta	aaaaaaaaaa	aaaaaaaaaa				2189

&lt;210&gt; 117

&lt;211&gt; 1763

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 117

ggtggcctag	agatgctgct	gccgcggttg	cagttgtcgc	gcacgcctct	gcccgccagc	60
ccgctccacc	gccgtagcgc	ccgagtgtcg	gggggagcac	ccgagtcggg	ccatgaggcc	120
gggaaccgcg	ctacaggccg	tgctgctggc	cgtgctgctg	gtggggctgc	ggggccgcac	180
gggtcgctg	ctgagtgggc	agccagtctg	ccggggaggg	acacagaggg	cttgttataa	240
agtcatttac	ttccatgata	cttctcgaag	actgaacttt	gaggaagcca	aagaagcctg	300
caggagggga	tggaggccag	ctagtcagca	tcgagttctg	aagatgaaca	gaaactgata	360
gaaaakttca	ttgaaaacct	cttgccatct	gatggtgayt	tctggawtgg	gctcaggagg	420
cgtgaggaga	aacaaagcaa	tagcacagct	gccaggacct	ttatgcttgg	actgatggca	480
gcataatcaca	atttaggaac	tggtatgtgg	atgagccgtc	ctgcggcagc	gaggtctgcg	540
tggtcatgta	ccatcagcca	tcggcacccg	ctggcatcgg	agggccctac	atgttccagt	600
ggaatgatga	ccggtgcaac	atgaagaaca	atttcatttg	caaataattct	gatgagaaac	660
cagcagttcc	ttctagagaa	gctgaagggt	aggaaacaga	gctgacaaca	cctgtacttc	720
cagaagaaaac	acaggaagaa	gatgccaaaa	aaacatttaa	agaaagtaga	gaagctgcct	780
tgaatctggc	ctacatccta	atccccagca	ttccccctct	cctcctcctt	gkggtcacca	840
cagttgtatg	ttgggttttg	atctgtagaa	aaagaaaacg	ggagcagcca	gaccctagca	900
caaagaagca	acacaccatc	tggccctctc	ctcaccaggg	aaacagcccg	gaccctagagg	960
tctacaatgt	catwagaaaa	caaagcgaag	ctgacttagc	tgagaccggg	ccagacctga	1020
agaatatattc	attccgagtg	tgttcggggg	aagccactcc	cgatgacatg	tcttgtgact	1080
atgacaacat	ggctgtgaac	ccatcagaaa	gtgggtttgt	gactctgggtg	agcgtggaga	1140
gtggatttgt	gaccaatgac	atztatgagt	tctccccaga	ccaaatgggg	aggagtaagg	1200
agtctggatg	gggtggaaaat	gaaatatatg	gttatttaga	catataaaaa	actgaaactg	1260
acaacaatgg	aaaagaaatg	ataagcaaaa	tcctcttatt	ttctataagg	aaaatacaca	1320
gaaggtctat	gaacaagctt	agatcaggtc	ctgtggatga	gcattgtggtc	cccacgacct	1380
cctgttggac	ccccacgttt	tggctgtatc	ctttatccca	gccagtcate	cagctcgacc	1440
ttatgagaag	gtaccttgcc	caggtctggc	acatagtaga	gtctcaataa	atgtcacttg	1500
gttggttgta	tctaactttt	aagggaacaga	gctttacctg	gcagtataaa	agatgggctg	1560
tggagcttgg	aaaaccacct	ctgttttcct	tgctctatac	agcagcacat	attatcatac	1620
agacagaaaa	tccagaatct	tttcaaagcc	cacatatggt	agcacagggt	ggcctgtgca	1680

tcggcaattc tcatatctgt ttttttcaaa gaataaaatc aaataaagag caggaaaaaa 1740  
 aaaaaaaaaa aaaaaaactc gag 1763

<210> 118  
 <211> 1375  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (18)  
 <223> n equals a,t,g, or c

<400> 118  
 ggcgggcccgc gcggaagngg cggckgcgcg gccggggcag ccatgtcgcc attgtctgcg 60  
 gcgcgggcccgc ccctgcgggt ctacgcggta ggcgcgcgcg tgatcctggc gcagctgctg 120  
 cggcgctgcc gsgggggctt cctggagcca gttytcccc cagcacctga ccgtgtcgt 180  
 atagtgcagg gagggacaga tggcattggc tattctacag cgaacatctg gcgagacttg 240  
 gcatgcatgt tatcatagct ggaaataatg acagcaaagc caaacaagtt gtaagcaaaa 300  
 taaaagaaga aaccttgaac gacaaagtgg aatttttata ctgtgacttg gcttccatga 360  
 cttccatccg gcagtttgtg cagaagttca agatgaagaa gattcctctc catgtcctga 420  
 tcaacaatgc tggggtgatg atggtcctc agaggaaaac cagagatgga ttcgaagaac 480  
 atttcggcct gaactaccta gggcaacttc tgctgaccaa ccttctcttg gatacgtga 540  
 aagagtctgg gtcccctggc cacagtgcga ggggtggcac cgtctcctct gccacccatt 600  
 acgtcgctga gctgaacatg gatgacctc agagcagtg cgtctactca cccacgcag 660  
 cctacgcccga gagcaagctg gcccttgctc tggttcaccta ccacctccag cggctgctgg 720  
 cggctgaggg aagccacgtg accgccaacg tgggtggacc cgggggtggc aacacggacs 780  
 tctacaagca cgtgttctg gccaccgcgc tggcgaagaa gcttctcggc tgggtgcttt 840  
 tcaagacccc cgatgaagga gcgtggactt ccatctacgc agcagtcacc ccagagctgg 900  
 aaggagtgg tggccgttac ctatataacg agaaagagac caagtccctc cagtcacct 960  
 acaaccagaa actgcagcag cagctgtggc ctaagagttg tgagatgact ggggtccttg 1020  
 atgtgacctt gtgatatcct gtctcaggat agctgctgcc ccaagaaaca cattgcacct 1080  
 gccaatagct tgtgggtctg tgaagactgc ggtgtttgag tttctcacac ccacctgcc 1140  
 acagggctct gtcctctagt tttgagacag ctgcctcaac ctctgcagaa cttcaagaag 1200  
 ccaataaaac attttgagg ataatacacc caagtggctt tcaaccataa actttgtgat 1260  
 tccaaagtgc ccagttgtca caggtgccat aaataattac attttccaac ataaaaaaaa 1320  
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaagggcg gccgc 1375

<210> 119  
 <211> 1022  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (937)  
 <223> n equals a,t,g, or c

<220>  
 <221> SITE  
 <222> (990)  
 <223> n equals a,t,g, or c

<400> 119  
 ggcacgagag agccrgcggc ccgagccggg gccatggcga actgctgagc tgcgtcctag 60  
 gcccccggt ctacaaaatc taccgggaga gggactctga aagggcccg gccagcgtcc 120  
 ctgagacgcc aacggcagtc actgcccccc attccagctc ctgggatacg tactatcagc 180

cccgtgccct	ggagaaacat	gctgacagca	tcctggcact	ggcttcagta	ttctgggtcca	240
tctcttatta	ctcctctccc	ttcgcccttct	tctacttgta	caggaaagggt	tacttgagtt	300
tgtccaaagt	ggtgccgttt	tctcactatg	ctgggacatt	gctgctactt	ctggcagggtg	360
tggcctgctc	cgaggcattg	gccgctggac	caacccccag	taccggcagt	tcaccacccat	420
cttggaagca	acacatcgga	accagtcttc	agaaaacaag	aggcagcttg	ccaactacaa	480
ctttgacttc	cggagctggc	cagtcgactt	ccactgggaa	gaacccagca	gccggaagga	540
gtctcgaggg	ggcccttccc	gccgggggtgt	ggccctgctt	cgcccagagc	ccctgcaccg	600
ggggacagca	gacacccctc	tcaaccgggt	taagaagctg	ccttgctcaga	tcaccagcta	660
cctggtggcg	cacacccctag	ggcgccggat	gctgtatcca	ggctctgtgt	acctgctgca	720
gaaggccctc	atgcctgtgc	tgctgcaggg	ccaggcccca	ctggtggaag	agtgtaatgg	780
gcgcggggca	aagctgctgg	cctgtgatgg	caatgagatt	gacaccatgt	ttgtggaccg	840
gcgggggaca	gctgagcccc	aggacagaag	ctggtgatct	gctgtgargg	gaatgctggg	900
ttttatgarg	tgggctgcgt	ctccamgccc	ctggaanctg	gatattcatc	ckgggtggaa	960
tcaccagct	ttgctggaac	acagggggtn	ccattccccg	aaaatgaagg	ctaatgccat	1020
gg						1022

&lt;210&gt; 120

&lt;211&gt; 2311

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (654)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (2293)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 120

tcgaccacg	cgctccgcttc	ctcatgctgc	ccgctcgcct	accgttcagg	ctgctgagcc	60
ttttccttcg	tggatccgct	cccacggcag	cgcgccatgg	cctccgggag	ccgctcctgg	120
agaggagggtg	cgctgccgcc	tcctccttec	agcactcctc	gagctctggga	cgcgagcttc	180
cttatgaccc	cgtggacacg	gagggctttg	gagaagggtg	tgacatgcag	gagcgttttc	240
tggtcccggga	gtacatcctg	gatccggagc	cgcaaccac	ccgcgaaaag	cagctgcagg	300
agctccagca	acagcaggag	gaggaggagc	gacagaggca	gcagcggcgg	gaggagcggc	360
gacagcaaaa	cctacggggc	aggtcccggg	agcaccgggt	cgtggggcac	ccggaccggg	420
cattgccgcc	cagcggcggtg	aactgctcgg	gctgcggggc	asascctgcac	tgccaggacg	480
ccggagtgcc	cggctacctg	ccccgagaga	agttcctccg	cacggcggag	gcagacggcg	540
ggctggcacg	gaccgtgtgc	cagcgtgtgt	ggctgctgtc	gcaccaccgg	cgcgctctac	600
gcctgcagggt	gagccgcgag	cagtacctgg	agctgggtgag	cgccgcgttg	cgngggcccg	660
gccccctcct	ggtgctctac	atgggtggacc	tgctggacct	gcccgcgcgc	ctgctgcccc	720
acttgcccgc	gctgggtggg	cccaagcagc	tgatcgtgct	gggaaacaaa	gtggaccttc	780
tgccccagga	tgctcctggc	taccggcaga	ggctgcggga	gcgactgtgg	gaggactgtg	840
cccgcgcggg	gctcctgctg	gccccctggca	ccaagggcca	cagcgcggcg	tcaaggacga	900
gccacaggac	ggggagaatc	cgaatccgcc	gaactgggtcc	cgcacagtgg	tcagggacgt	960
gcggctgac	agcgccaaga	ccggctatgg	agtgggaagag	ttgatctctg	cccttcagcg	1020
ctcctggcgc	taccgtgggg	acgtctactt	agtgggcgcc	accaacgcgc	gcaaataccac	1080
tctctttaac	acgtcctctg	agtcggatta	ctgcactgcc	aagggtcccg	aggccatcga	1140
cagagccacc	atctcccttc	ggccagggtac	tacattaaac	cttctgaagt	ttcctatttg	1200
caacccaact	ccttacagaa	tgtttaaaag	gcatcaaaga	cttaaaaaag	attcaactca	1260
agctgaagaa	gatcttagtg	agcaagaaca	aaatcagctt	aatgtcctca	aaaagcatgg	1320
ttatgtcgtg	ggaagagttg	gaaggacatt	cttgattatc	gaagaacaga	aggataacat	1380
tccttttgag	tttgatgctg	attcacttgc	ctttgacatg	gaaaatgacc	ctggttatggg	1440
tacacacaaa	tccaccaaac	aagtagaatt	gactgcacaa	gatgtgaaag	atgcccatg	1500
gttttatgac	acccttgga	ttacaaaaga	aaattgtatt	ttaaatcttc	taacagaaaa	1560

agaagtaaat	attgtttttg	caacacagtc	cattgtttcca	agaacttttg	tgcttaaacc	1620
aggaatggtt	ctgttttttg	gtgctatagg	ccgcatagat	ttcctgcagg	gaaatcagtc	1680
agcttggttt	acagtcgtgg	cttccaacat	cctccctgtg	catatcacct	ccttggacag	1740
ggcagacgct	ctgtatcaga	agcatgcagg	tcatacgta	ctccagattc	caatgggtgg	1800
aaaagaacga	atggcrrgat	ttcctcctct	tgttgctgaa	gacattatgt	taaaagaagg	1860
actgggggca	tctgaagcag	tggccgacat	caagttttcc	tctgcagggt	gggtttcagt	1920
aacacctaat	tttaaggaca	gactgcatct	ccgaggctat	acacctgaag	gaacagtttt	1980
gaccgtccgg	ccccctctct	tgccatatat	tgtaaacatc	aaaggacagc	gcatcaagaa	2040
aagtgtggcc	tataaaacca	agaagcctcc	ttcccttatg	tacaacgtga	ggaagaagaa	2100
aggaaagata	aatgtatgag	accgaccttg	ttcactccag	atattaactg	tattgaacac	2160
aacaaaatac	attgaatttg	tattaaacat	ataacgcata	aataaagctc	ccattcttac	2220
ccttaaaaaa	aaaaaaaaag	ggcggccgct	ctagaggatc	caagcttacg	tacgcgtgca	2280
tgcgacgtca	tanctcgtct	ataggaactg	g			2311

<210> 121  
 <211> 1286  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (1284)  
 <223> n equals a,t,g, or c

<400> 121	
gggcgcgcgg	gtgaaaggcg
agacgctgac	cacgttcctc
gcccggagcc	atgcgacccc
gctgctcctg	ctgctgcagc
gcaaaaggcg	cactccggca
gggccagcag	gagtgccctg
cctgggatac	caggtcggga
tttgaggagt	cctggacacc
atagatcttg	ggaaaattgc
agagttttgt	tcagtggctc
tatttcacat	tcaatggagc
ttggaccaag	gaagccctga
gaaggacttt	gtgaaggaat
tgttcagatt	acccaaaagg
attgaagaac	tacccaaaata
ccttggaatg	gttcacttaa
agcaaagcta	aatatgttta
ttattcattt	tgcttcaatc
ttcttcatag	tcacattctc
ttctcttagt	atagcatttt
aatgttaaga	atttttttta
aaaaaaaaaa	aaaaaaaaaa
aaanaa	

<210> 122  
 <211> 1380  
 <212> DNA  
 <213> Homo sapiens

<400> 122	
cagacccgcg	gggcaaacgg
gggcccgcgc	aggcgccgc
tctgcctct	agccatgggg
actggggcca	agagccggga
gcgcgggcgc	aaaggcacca
tggggggttc	tgccggcctt
gggcctggtg	ctgtgcctgg

tgggctgggg	gggtctgac	ctggcggtg	ggctgccc	gtggcaggt	accgccttcc	240
tggaccacaa	catcgtgac	gcgcagacca	cctggaaggg	gctgtggatg	tcgtgctgtg	300
tgcagagcac	gggcacatgc	agtgc aaagt	gtacgactcg	gtgctggctc	tgagcaccga	360
ggtgcaggcg	gcgcggggcg	tcaccgtgag	cgccgtgctg	ctggcgttcg	ttgcgctctt	420
cgtgaccctg	gcggggcgcg	agtgcaccac	ctgcgtggcc	ccggggcccg	ccaaggcgcg	480
tgtggccctc	acgggaggcg	tgctctacct	gttttgcggg	ctgctggcgc	tcgtgccact	540
ctgctggttc	gccaacattg	tcgtccgcga	gttttacgac	ccgtctgtgc	ccgtgtcgca	600
gaagtacgag	ctggggcgac	gctgtacatc	ggctggggcg	ccaccgcgct	gctcatggta	660
ggcggctgcc	tcttgctgct	cgggcgctgg	gtctgcaccg	gccgtcccg	cctcagcttc	720
cccgtgaagt	actcagcgcc	gcggcgggcc	acggccaccg	gcgactacga	caagaagaac	780
tacgtctgag	ggcgctgggc	acggccgggc	ccctcctgcc	agccacgcct	gcgaggcggt	840
ggataagcct	ggggagcccc	gcattggaccg	cggttccgc	cggttagcgc	ggcgcgccag	900
ctcctcgga	cgccggctc	tgcccccga	cgccgctcct	ggatccgctc	ctgcctgcgc	960
ccgcagctga	ccttctcctg	ccactagccc	ggccctgccc	ttaacagacg	gaatgaagtt	1020
tccttttctg	tgccggcgcg	tgtttcata	ggcagagcgg	gtgtcagact	gaggatttgc	1080
cttccctcc	aagacgctgg	gggtcttggc	tgctgcctta	cttcccagag	gctcctgctg	1140
acttcggagg	ggcggatgca	gagcccagg	ccccaccgg	aagatgtgta	cacctggtct	1200
ttactccatc	ggcagggccc	gagcccagg	accagtgact	tggcctggac	ctcccgggtct	1260
cactccagca	tctccccagg	caaggcttgt	gggcaccgga	gcttgagaga	gggcgggagt	1320
gggaaggcta	agaatctgct	tagtaaatgg	tttgaactct	caaaaaaaaa	aaaaaaaaaa	1380

&lt;210&gt; 123

&lt;211&gt; 3793

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (1102)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (1132)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (1199)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (1228)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (1229)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (1231)

&lt;223&gt; n equals a,t,g, or c

&lt;220&gt;

&lt;221&gt; SITE



&lt;222&gt; (3176)

&lt;223&gt; n equals a,t,g, or c

&lt;400&gt; 123

ggggcttcat	acaggaaatc	tattgctgtg	tcaagttcca	gagaaaagct	tctgttcgtc	60
caagttacta	accaggctaa	accacataga	cgtgaaggaa	ggggctagaa	ggaagggagt	120
gccccactgt	tgatggggta	agaggatcct	gtactgagaa	gttgaccaga	gagggctca	180
ccatgcgcac	agttccttct	gtaccagtgt	ggaggaaaag	tactgagtga	agggcagaaa	240
aagagaaaac	agaaatgctc	tgcccttgga	gaactgctaa	cctagggcta	ctgttgattt	300
tgactatctt	cttagtgggc	gaagcggagg	gtgctgctca	accaaacaac	tcattaatgc	360
tgcaaaactag	caaggagaat	catgcttttag	cttcaagcag	tttatgtatg	gatgaaaaac	420
agattacaca	gaactactcg	aaagtactcg	cagaagttaa	cacttcatgg	cctgtaaaaga	480
tggctacaaa	tgctgtgctt	tggtgcccctc	ctatcgcatt	aagaaatttg	atcataataa	540
catggggaat	aattctgaga	ggccagcctt	cctgcacaaa	agcctacaag	aaagaaacaa	600
atgagaccaa	ggaaaccaac	tgtactgatg	agagaataac	ctgggtctcc	agaccgatc	660
agaattcggga	ccttcagatt	cgtaccgtgg	ccatcactca	tgacgggtat	tacagatgca	720
taatggtaac	acctgatggg	aatttccatc	gtggatatca	cctccaagtg	ttagttacac	780
ctgaagtgc	cctgtttcaa	aacaggaata	gaactgcagt	atgcaaggca	gttgcaggga	840
agccagctgc	gcatactctc	tggatcccag	agggcgattg	tgccactaag	caagaatact	900
ggagcaatgg	cacagtgcct	gttaagagta	catgccactg	ggaggtccac	aatgtgtcta	960
cctggaactg	ccacgtctcc	catttgactg	gcaacaagag	tctgtacata	gagctacttc	1020
ctgttccagg	tgccaaaaaa	tcatacaaat	tatatattcc	atatatcatc	cttactatta	1080
ttattttgac	catcgtggga	tncattttggt	tggtgaaagt	caatggctgc	anaaaatata	1140
aattgaataa	accagaatct	actccagttg	ttgaggagga	tgaaatgcag	ccctatgcnt	1200
tttacacaga	gaagaacaat	cctctctnng	ntactacaaa	caaggtgaag	gcattctgagg	1260
cattacaaaag	tgaagttgac	acagacctcc	atactttata	agttgttgga	ctctagtacc	1320
aagaaacaac	aacaaacgag	atacattata	attactgtct	gattttctta	cagttctaga	1380
atgaagactt	atattgaaat	taggttttcc	aaggttctta	gaagacattt	taattggattc	1440
tcattcatac	ccttgtataa	ttggaatttt	tgattcttag	ctgctaccag	ctagttctct	1500
gaagaactga	tggtattaca	aagaaaatac	atgcccata	ccaaatattc	aaattgtgca	1560
ggacagtaaa	taatgaaaac	caaatttcc	caagaaataa	ctgaagaagg	agcaagtgtg	1620
aacagtttct	tgtgtatcct	ttcagaatat	tttaatgtac	atatgacatg	tgtatatgcc	1680
tatgggtatat	gtgtcaattt	atgtgtcccc	ttacatatac	catgcaccct	atctttgtca	1740
aggcaccagt	gggaacaata	cactgcatta	ctgttctata	catatgaaaa	cctaataata	1800
taagtcttag	agatcatttt	atatcatgac	aagttagagct	acctcattct	ttttaatgggt	1860
tatataaaat	tccattgtat	agttatatca	ttattttaatt	aaaaacaacc	ctaattgatgg	1920
atatttagat	tcttttaagt	tttgtttatt	tcttttaagt	tttgtttgtg	gtataaacia	1980
taccacatag	aatgtttctt	gtgcatatat	ctctttgttt	ttgagtatat	ctgtaggata	2040
actttcttga	gtggaattgt	caggtcaaag	ggtttgtgca	ttttactatt	gatataatg	2100
ttaaatttgt	tcaaataatat	atgtcaaatt	ccctccaaca	ttgtttaaat	gtgcctttcc	2160
ctaaatttct	attttaataa	ctgtactatt	cctgcttcta	cagttgccac	tttctctttt	2220
taatcaacca	gattaaatat	gatgtgagat	tataataaga	attatactat	tttaataaaaa	2280
tggatttata	tttttggtca	tgtttgtaag	agagtgaatg	cacgtgtgag	aacattagct	2340
tcttctgaac	tcatttatatc	tccacagagg	tggtgatact	tgatgcctaa	cagttttgca	2400
gatgtgctac	attggaattg	tgtattttta	tggtgtacat	tctattgtga	tatatattatt	2460
gaataattaa	tgtctattga	ccatataagt	ggcgaaaaat	gcaccataga	ggacatgggg	2520
tattttatta	caaactatga	gctacataat	aagcaagtgg	ccatgggatg	gcattgaccct	2580
cccctccata	tttttggtga	gcaaaatatt	ggcaatgttt	atgtaaatca	ttgttaatat	2640
catgaaatta	tttttaatta	aaaacataag	tctatttgct	ccatagcaga	aaaaacatga	2700
gaagtttttt	catcatgata	gaaattgaaa	caaaatatat	tcattcttca	atcataccat	2760
ctgagatttt	taagacagct	attttgtctt	ataagtatat	ttttctccct	ctagacattt	2820
cagttactat	ggattttgtc	ctcaaaggga	cytttagtct	aattttggga	tgtaaagcta	2880
atcttaatga	cacttggcac	atgataattt	gatcaagcca	ttttgacttg	acaaaaaagc	2940
agtgtccatt	aggtttctgc	atataaatat	taccaagcaa	tgttcacaat	agacatcatt	3000
acactgtcct	tgaaatttat	taattcttca	tccaacctga	gttgagctga	ggctcatagt	3060
taggttcaag	actatctgtt	taaatattac	tgaaaaacaa	agtaagacag	tactatgctt	3120
acctcttaac	ttgataatgt	caaaccaggc	atgttaaatg	acatcataga	aaaganttca	3180
agataattta	tagaagttaa	attatattgt	acagaaaata	attgtatgaa	aatctctact	3240
atggggctgg	aacatgggtg	aacattagaa	tgatataaaa	aattatatat	attctccaaa	3300

```

tccacgctag acctgtcaaa ttagagaatc tagagattag acctggcgtg tcagcaaggt 3360
catccaggaa gcagaggctg agacggagtt aggtgtgatt acttacatag tcgattacat 3420
tttacaata acattttata tgtctcattt actgtgcttt ctcccatcc cattttgtat 3480
cttttccttt gctttgctag atttgtcaat tttctctctc tttctgtctc tctctctttc 3540
aatatctcta ataatttgaa agtaattcat cataactaaa tatctattgg gggttatgctt 3600
cacttacaaa cttctgaaaa cggctttact gagatataat tgatatattt aagtgtacag 3660
tttgttaaat tttgcacata tttaaaatgt ggactttggt aaatggtgac atagttttac 3720
atctgtgaaa ccatacagcat aatcaagata ataaacttgt ccataccccc ccaaaaaaaaa 3780
aaaaaaaaaa aaa 3793

```

<210> 124  
 <211> 370  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (370)  
 <223> Xaa equals stop translation

<400> 124

Met Leu Gly Ala Phe Val Trp Pro Ser Leu Leu Leu Leu Ala Ala Ala  
 1 5 10 15

Cys Ile Cys Leu Leu Thr Phe Ile Asn Cys Ala Tyr Val Lys Trp Gly  
 20 25 30

Thr Leu Val Gln Asp Ile Phe Thr Tyr Ala Lys Val Leu Ala Leu Ile  
 35 40 45

Ala Val Ile Val Ala Gly Ile Val Arg Leu Gly Gln Gly Ala Ser Thr  
 50 55 60

His Phe Glu Asn Ser Phe Glu Gly Ser Ser Phe Ala Val Gly Asp Ile  
 65 70 75 80

Ala Leu Ala Leu Tyr Ser Ala Leu Phe Ser Tyr Ser Gly Trp Asp Thr  
 85 90 95

Leu Asn Tyr Val Thr Glu Glu Ile Lys Asn Pro Glu Arg Asn Leu Pro  
 100 105 110

Leu Ser Ile Gly Ile Ser Met Pro Ile Val Thr Ile Ile Tyr Ile Leu  
 115 120 125

Thr Asn Val Ala Tyr Tyr Thr Val Leu Asp Met Arg Asp Ile Leu Ala  
 130 135 140

Ser Asp Ala Val Ala Val Thr Phe Ala Asp Gln Ile Phe Gly Ile Phe  
 145 150 155 160

Asn Trp Ile Ile Pro Leu Ser Val Ala Leu Ser Cys Phe Gly Gly Leu  
 165 170 175

Asn Ala Ser Ile Val Ala Ala Ser Arg Leu Phe Phe Val Gly Ser Arg  
 180 185 190

Glu Gly His Leu Pro Asp Ala Ile Cys Met Ile His Val Glu Arg Phe

195	200	205
Thr Pro Val Pro Ser Leu Leu Phe Asn Gly Ile Met Ala Leu Ile Tyr		
210	215	220
Leu Cys Val Glu Asp Ile Phe Gln Leu Ile Asn Tyr Tyr Ser Phe Ser		
225	230	235 240
Tyr Trp Phe Phe Val Gly Leu Ser Ile Val Gly Gln Leu Tyr Leu Arg		
	245	250 255
Trp Lys Glu Pro Asp Arg Pro Arg Pro Leu Lys Leu Ser Val Phe Phe		
	260	265 270
Pro Ile Val Phe Cys Leu Cys Thr Ile Phe Leu Val Ala Val Pro Leu		
	275	280 285
Tyr Ser Asp Thr Ile Asn Ser Leu Ile Gly Ile Ala Ile Ala Leu Ser		
	290	295 300
Gly Leu Pro Phe Tyr Phe Leu Ile Ile Arg Val Pro Glu His Lys Arg		
	305	310 315 320
Pro Leu Tyr Leu Arg Arg Ser Trp Gly Leu Pro Gln Gly Thr Ser Arg		
	325	330 335
Ser Cys Val Cys Gln Leu Leu Gln Lys Trp Ile Trp Lys Met Glu Glu		
	340	345 350
Arg Cys Pro Ser Asn Gly Ile Pro Ser Leu Thr Lys His His Leu Glu		
	355	360 365
Ser Xaa		
370		

---

<210> 125  
 <211> 86  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (86)  
 <223> Xaa equals stop translation

<400> 125  
 Met Gly Phe Trp Cys Gly Cys Pro Phe Cys Leu Leu Val Val Leu Leu  
 1 5 10 15  
 Thr Asp Arg Thr Leu Ser Cys Arg Ser Val Gly Val Pro Cys Asn Val  
 20 25 30  
 Arg Cys Gln Cys Ala Pro Ala Gly Gly Cys Leu Pro Val Arg Leu Leu  
 35 40 45  
 Ala Gly Gln Gly Ser Gly Thr His Leu Arg Arg Gln Ser Ala Arg Ser  
 50 55 60

Gln Ile Ser Ser Cys Met Leu Gly Glu Pro Leu Leu Ser Ser Lys Leu  
 65 70 75 80

Ser Asp Arg Asp Ile Xaa  
 85

<210> 126

<211> 44

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (44)

<223> Xaa equals stop translation

<400> 126

Met Tyr Thr Lys Thr His Lys Phe Lys Phe Tyr Asn Phe Leu Ser Leu  
 1 5 10 15

Trp Ile Trp Lys Ile Phe Phe Leu Leu Phe Phe Ile Leu Ile Val Ala  
 20 25 30

Leu Ala Phe Pro Ile Pro Cys Leu Ser Ile Phe Xaa  
 35 40

<210> 127

<211> 319

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (264)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (303)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 127

Met Asn Thr Asp His Leu Arg Leu Thr Val Pro Asn Gly Ile Gly Ala  
 1 5 10 15

Leu Lys Leu Arg Glu Met Glu His Tyr Phe Ser Gln Gly Leu Ser Val  
 20 25 30

Gln Leu Phe Asn Asp Gly Ser Lys Gly Lys Leu Asn His Leu Cys Gly  
 35 40 45

Ala Asp Phe Val Lys Ser His Gln Lys Pro Pro Gln Gly Met Glu Ile  
 50 55 60

Lys Ser Asn Glu Arg Cys Cys Ser Phe Asp Gly Asp Ala Asp Arg Ile  
 65 70 75 80

Val	Tyr	Tyr	Tyr	His	Asp	Ala	Asp	Gly	His	Phe	His	Leu	Ile	Asp	Gly	85	90	95
Asp	Lys	Ile	Ala	Thr	Leu	Ile	Ser	Ser	Phe	Leu	Lys	Glu	Leu	Leu	Val	100	105	110
Glu	Ile	Gly	Glu	Ser	Leu	Asn	Ile	Gly	Val	Val	Gln	Thr	Ala	Tyr	Ala	115	120	125
Asn	Gly	Ser	Ser	Thr	Arg	Tyr	Leu	Glu	Glu	Val	Met	Lys	Val	Pro	Val	130	135	140
Tyr	Cys	Thr	Lys	Thr	Gly	Val	Lys	His	Leu	His	His	Lys	Ala	Gln	Glu	145	150	155
Phe	Asp	Ile	Gly	Val	Tyr	Phe	Glu	Ala	Asn	Gly	His	Gly	Thr	Ala	Leu	165	170	175
Phe	Ser	Thr	Ala	Val	Glu	Met	Lys	Ile	Lys	Gln	Ser	Ala	Glu	Gln	Leu	180	185	190
Glu	Asp	Lys	Lys	Arg	Lys	Ala	Ala	Lys	Met	Leu	Glu	Asn	Ile	Ile	Asp	195	200	205
Leu	Phe	Asn	Gln	Ala	Ala	Gly	Asp	Ala	Ile	Ser	Asp	Met	Leu	Val	Ile	210	215	220
Glu	Ala	Ile	Leu	Ala	Leu	Lys	Gly	Leu	Thr	Val	Gln	Gln	Trp	Asp	Ala	225	230	235
Leu	Tyr	Thr	Asp	Leu	Pro	Asn	Arg	Gln	Leu	Lys	Val	Gln	Val	Ala	Asp	245	250	255
Arg	Arg	Val	Ile	Ser	Thr	Thr	Xaa	Ala	Glu	Arg	Gln	Ala	Val	Thr	Pro	260	265	270
Pro	Gly	Leu	Gln	Glu	Ala	Ile	Asn	Asp	Leu	Val	Lys	Lys	Tyr	Lys	Leu	275	280	285
Ser	Arg	Ala	Phe	Val	Arg	Pro	Ser	Gly	Thr	Glu	Asp	Val	Val	Xaa	Ser	290	295	300
Ile	Cys	Arg	Ser	Arg	Leu	Thr	Arg	Lys	Cys	Arg	Ser	Pro	Cys	Thr		305	310	315

<210> 128  
 <211> 46  
 <212> PRT  
 <213> Homo sapiens  
 <220>  
 <221> SITE  
 <222> (46)  
 <223> Xaa equals stop translation  
 <400> 128  
 Met Asp Met Val Cys Phe Cys Ile Tyr Leu Gly Leu Leu Lys Phe Ile  
 1 5 10 15

Ser Ala Ile Phe Cys Ser Phe Ser Glu Glu Val Leu Tyr Ile Ser Phe  
                   20                                  25                                  30

Val Lys Cys Ile Pro Lys Tyr Phe Val Glu Met Leu Leu Xaa  
                   35                                  40                                  45

<210> 129

<211> 709

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (189)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (275)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (414)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (438)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (641)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (643)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (696)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (697)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 129

Met Ala Gly Leu Asn Cys Gly Val Ser Ile Ala Leu Leu Gly Val Leu  
   1                                  5                                  10                                  15

Leu Leu Gly Ala Ala Arg Leu Pro Arg Gly Ala Glu Ala Phe Glu Ile  
                   20                                  25                                  30

Ala Leu Pro Arg Glu Ser Asn Ile Thr Val Leu Ile Lys Leu Gly Thr  
           35                          40                          45  
 Pro Thr Leu Leu Ala Lys Pro Cys Tyr Ile Val Ile Ser Lys Arg His  
           50                          55                          60  
 Ile Thr Met Leu Ser Ile Lys Ser Gly Glu Arg Ile Val Phe Thr Phe  
           65                          70                          75                          80  
 Ser Cys Gln Ser Pro Glu Asn His Phe Val Ile Glu Ile Gln Lys Asn  
                           85                          90                          95  
 Ile Asp Cys Met Ser Gly Pro Cys Pro Phe Gly Glu Val Gln Leu Gln  
                           100                          105                          110  
 Pro Ser Thr Ser Leu Leu Pro Thr Leu Asn Arg Thr Phe Ile Trp Asp  
                           115                          120                          125  
 Val Lys Ala His Lys Ser Ile Gly Leu Glu Leu Gln Phe Ser Ile Pro  
                           130                          135                          140  
 Arg Leu Arg Gln Ile Gly Pro Gly Glu Ser Cys Pro Asp Gly Val Thr  
                           145                          150                          155                          160  
 His Ser Ile Ser Gly Arg Ile Asp Ala Thr Val Val Arg Ile Gly Thr  
                           165                          170                          175  
 Phe Cys Ser Asn Gly Thr Val Ser Arg Ile Lys Met Xaa Glu Gly Val  
                           180                          185                          190  
 Lys Met Ala Leu His Leu Pro Trp Phe His Pro Arg Asn Val Ser Gly  
                           195                          200                          205  
 Phe Ser Ile Ala Asn Arg Ser Ser Ile Lys Arg Leu Cys Ile Ile Glu  
                           210                          215                          220

---

Ser Val Phe Glu Gly Glu Gly Ser Ala Thr Leu Met Ser Ala Asn Tyr  
                           225                          230                          235                          240  
 Pro Glu Gly Phe Pro Glu Asp Glu Leu Met Thr Trp Gln Phe Val Val  
                           245                          250                          255  
 Pro Ala His Leu Arg Ala Ser Val Ser Phe Leu Asn Phe Asn Leu Ser  
                           260                          265                          270  
 Asn Cys Xaa Arg Lys Glu Glu Arg Val Glu Tyr Tyr Ile Pro Gly Ser  
                           275                          280                          285  
 Thr Thr Asn Pro Glu Val Phe Lys Leu Glu Asp Lys Gln Pro Gly Asn  
                           290                          295                          300  
 Met Ala Gly Asn Phe Asn Leu Ser Leu Gln Gly Cys Asp Gln Asp Ala  
                           305                          310                          315                          320  
 Gln Ser Pro Gly Ile Leu Arg Leu Gln Phe Gln Val Leu Val Gln His  
                           325                          330                          335  
 Pro Gln Asn Glu Ser Asn Lys Ile Tyr Val Val Asp Leu Ser Asn Glu

340	345	350
Arg Ala Met Ser Leu Thr Ile Glu Pro Arg Pro Val Lys Gln Ser Arg		
355	360	365
Lys Phe Val Pro Gly Cys Phe Val Cys Leu Glu Ser Arg Thr Cys Ser		
370	375	380
Ser Asn Leu Thr Leu Thr Ser Gly Ser Lys His Lys Ile Ser Phe Leu		
385	390	400
Cys Asp Asp Leu Thr Arg Leu Trp Met Asn Val Glu Lys Xaa Ile Ser		
405	410	415
Cys Thr Asp His Arg Tyr Cys Gln Arg Lys Ser Tyr Ser Leu Gln Val		
420	425	430
Pro Ser Asp Ile Leu Xaa Leu Pro Val Glu Leu His Asp Phe Ser Trp		
435	440	445
Lys Leu Leu Val Pro Lys Asp Arg Leu Ser Leu Val Leu Val Pro Ala		
450	455	460
Gln Lys Leu Gln Gln His Thr His Glu Lys Pro Cys Asn Thr Ser Phe		
465	470	475
Ser Tyr Leu Val Ala Ser Ala Ile Pro Ser Gln Asp Leu Tyr Phe Gly		
485	490	495
Ser Phe Cys Pro Gly Gly Ser Ile Lys Gln Ile Gln Val Lys Gln Asn		
500	505	510
Ile Ser Val Thr Leu Arg Thr Phe Ala Pro Ser Phe Arg Gln Glu Ala		
515	520	525
Ser Arg Gln Gly Leu Thr Val Ser Phe Ile Pro Tyr Phe Lys Glu Glu		
530	535	540
Gly Val Phe Thr Val Thr Pro Asp Thr Lys Ser Lys Val Tyr Leu Arg		
545	550	555
Thr Pro Asn Trp Asp Arg Gly Leu Pro Ser Leu Thr Ser Val Ser Trp		
565	570	575
Asn Ile Ser Val Pro Arg Asp Gln Val Ala Cys Leu Thr Phe Phe Lys		
580	585	590
Glu Arg Ser Gly Val Val Cys Gln Thr Gly Arg Ala Phe Met Ile Ile		
595	600	605
Gln Glu Gln Arg Thr Arg Ala Glu Glu Ile Phe Ser Leu Asp Glu Asp		
610	615	620
Val Leu Pro Lys Pro Ser Phe His His His Ser Phe Trp Val Asn Ile		
625	630	635
Xaa Asn Xaa Ser Pro Thr Ser Gly Lys Gln Leu Asp Leu Leu Phe Ser		
645	650	655



Val Thr Leu Thr Pro Arg Thr Val Asp Leu Thr Val Ile Leu Ile Ala  
660 665 670

Ala Val Gly Gly Gly Val Leu Leu Ser Ala Leu Gly Leu Ile Ile  
675 680 685

Cys Cys Val Lys Lys Lys Lys Xaa Xaa Thr Arg Gly Pro Ala Val Gly  
690 695 700

Ile Tyr Asn Gly Asn  
705

<210> 130

<211> 415

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (415)

<223> Xaa equals stop translation

<400> 130

Met Thr Lys Ala Arg Leu Phe Arg Leu Trp Leu Val Leu Gly Ser Val  
1 5 10 15

Phe Met Ile Leu Leu Ile Ile Val Tyr Trp Asp Ser Ala Gly Ala Ala  
20 25 30

His Phe Tyr Leu His Thr Ser Phe Ser Arg Pro His Thr Gly Pro Pro  
35 40 45

Leu Pro Thr Pro Gly Pro Asp Arg Asp Arg Glu Leu Thr Ala Asp Ser  
50 55 60

Asp Val Asp Glu Phe Leu Asp Lys Phe Leu Ser Ala Gly Val Lys Gln  
65 70 75 80

Ser Asp Leu Pro Arg Lys Glu Thr Glu Gln Pro Pro Ala Pro Gly Ser  
85 90 95

Met Glu Glu Asn Val Arg Gly Tyr Asp Trp Ser Pro Arg Asp Ala Arg  
100 105 110

Arg Ser Pro Asp Gln Gly Arg Gln Gln Ala Glu Arg Arg Ser Val Leu  
115 120 125

Arg Gly Phe Cys Ala Asn Ser Ser Leu Ala Phe Pro Thr Lys Glu Arg  
130 135 140

Ala Phe Asp Asp Ile Pro Asn Ser Glu Leu Ser His Leu Ile Val Asp  
145 150 155 160

Asp Arg His Gly Ala Ile Tyr Cys Tyr Val Pro Lys Val Ala Cys Thr  
165 170 175

Asn Trp Lys Arg Val Met Ile Val Leu Ser Gly Ser Leu Leu His Arg  
180 185 190

Gly Ala Pro Tyr Arg Asp Pro Leu Arg Ile Pro Arg Glu His Val His  
 195 200 205  
 Asn Ala Ser Ala His Leu Thr Phe Asn Lys Phe Trp Arg Arg Tyr Gly  
 210 215 220  
 Lys Leu Ser Arg His Leu Met Lys Val Lys Leu Lys Lys Tyr Thr Lys  
 225 230 235 240  
 Phe Leu Phe Val Arg Asp Pro Phe Val Arg Leu Ile Ser Ala Phe Arg  
 245 250 255  
 Ser Lys Phe Glu Leu Glu Asn Glu Glu Phe Tyr Arg Lys Phe Ala Val  
 260 265 270  
 Pro Met Leu Arg Leu Tyr Ala Asn His Thr Ser Leu Pro Ala Ser Ala  
 275 280 285  
 Arg Glu Ala Phe Arg Ala Gly Leu Lys Val Ser Phe Ala Asn Phe Ile  
 290 295 300  
 Gln Tyr Leu Leu Asp Pro His Thr Glu Lys Leu Ala Pro Phe Asn Glu  
 305 310 315 320  
 His Trp Arg Gln Val Tyr Arg Leu Cys His Pro Cys Gln Ile Asp Tyr  
 325 330 335  
 Asp Phe Val Gly Lys Leu Glu Thr Leu Asp Glu Asp Ala Ala Gln Leu  
 340 345 350  
 Leu Gln Leu Leu Gln Val Asp Arg Gln Leu Arg Phe Pro Pro Ser Tyr  
 355 360 365  
 Arg Asn Arg Thr Ala Ser Ser Trp Glu Glu Asp Trp Phe Ala Lys Ile  
 370 375 380  
 Pro Leu Ala Trp Arg Gln Gln Leu Tyr Lys Leu Tyr Glu Ala Asp Phe  
 385 390 395 400  
 Val Leu Phe Gly Tyr Pro Lys Pro Glu Asn Leu Leu Arg Asp Xaa  
 405 410 415

&lt;210&gt; 131

&lt;211&gt; 242

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 131

Met Gln Leu Gly Ser Val Leu Leu Thr Arg Cys Pro Phe Trp Gly Cys  
 1 5 10 15

Phe Ser Gln Leu Met Leu Tyr Ala Glu Arg Ala Glu Ala Arg Arg Lys  
 20 25 30

Pro Asp Ile Pro Val Pro Tyr Leu Tyr Phe Asp Met Gly Ala Ala Val  
 35 40 45

Leu Cys Ala Ser Phe Met Ser Phe Gly Val Lys Arg Arg Trp Phe Ala  
 50 55 60  
 Leu Gly Ala Ala Leu Gln Leu Ala Ile Ser Thr Tyr Ala Ala Tyr Ile  
 65 70 75 80  
 Gly Gly Tyr Val His Tyr Gly Asp Trp Leu Lys Val Arg Met Tyr Ser  
 85 90 95  
 Arg Thr Val Ala Ile Ile Gly Gly Phe Leu Val Leu Ala Ser Gly Ala  
 100 105 110  
 Gly Glu Leu Tyr Arg Arg Lys Pro Arg Ser Arg Ser Leu Gln Ser Thr  
 115 120 125  
 Gly Gln Val Phe Leu Gly Ile Tyr Leu Ile Cys Val Ala Tyr Ser Leu  
 130 135 140  
 Gln His Ser Lys Glu Asp Arg Leu Ala Tyr Leu Asn His Leu Pro Gly  
 145 150 155 160  
 Gly Glu Leu Met Ile Gln Leu Phe Phe Val Leu Tyr Gly Ile Leu Ala  
 165 170 175  
 Leu Ala Phe Leu Ser Gly Tyr Tyr Val Thr Leu Ala Ala Gln Ile Leu  
 180 185 190  
 Ala Val Leu Leu Pro Pro Val Met Leu Leu Ile Asp Gly Asn Val Ala  
 195 200 205  
 Tyr Trp His Asn Thr Arg Arg Val Glu Phe Trp Asn Gln Met Lys Leu  
 210 215 220  
 Leu Gly Glu Ser Val Gly Ile Phe Gly Thr Ala Val Ile Leu Ala Thr  
 225 230 235 240

---

Asp Gly

<210> 132  
 <211> 313  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (313)  
 <223> Xaa equals stop translation

<400> 132  
 Met Glu Ser Leu Tyr Asp Leu Trp Glu Phe Tyr Leu Pro Tyr Leu Tyr  
 1 5 10 15  
 Ser Cys Ile Ser Leu Met Gly Cys Leu Leu Leu Leu Cys Thr Pro  
 20 25 30  
 Val Gly Leu Ser Arg Met Phe Thr Val Met Gly His Leu Leu Val Lys  
 35 40 45

Pro Thr Ile Leu Glu Asp Leu Asp Glu Gln Ile Tyr Ile Ile Thr Leu  
 50 55 60  
 Glu Glu Glu Ala Leu Gln Arg Arg Leu Asn Gly Leu Ser Ser Ser Val  
 65 70 75 80  
 Glu Tyr Asn Ile Met Glu Leu Glu Gln Glu Leu Glu Asn Val Lys Thr  
 85 90 95  
 Leu Lys Thr Lys Leu Glu Arg Arg Lys Lys Ala Ser Ala Trp Glu Arg  
 100 105 110  
 Asn Leu Val Tyr Pro Ala Val Met Val Leu Leu Leu Ile Glu Thr Ser  
 115 120 125  
 Ile Ser Val Leu Leu Val Ala Cys Asn Ile Leu Cys Leu Leu Val Asp  
 130 135 140  
 Glu Thr Ala Met Pro Lys Gly Thr Arg Gly Pro Gly Ile Gly Asn Ala  
 145 150 155 160  
 Ser Leu Ser Thr Phe Gly Phe Val Gly Ala Ala Leu Glu Ile Ile Leu  
 165 170 175  
 Ile Phe Tyr Leu Met Val Ser Ser Val Val Gly Phe Tyr Ser Leu Arg  
 180 185 190  
 Phe Phe Gly Asn Phe Thr Pro Lys Lys Asp Asp Thr Thr Met Thr Lys  
 195 200 205  
 Ile Ile Gly Asn Cys Val Ser Ile Leu Val Leu Ser Ser Ala Leu Pro  
 210 215 220  
 Val Met Ser Arg Thr Leu Gly Ile Thr Arg Phe Asp Leu Leu Gly Asp  
 225 230 235 240  
 Phe Gly Arg Phe Asn Trp Leu Gly Asn Phe Tyr Ile Val Leu Ser Tyr  
 245 250 255  
 Asn Leu Leu Phe Ala Ile Val Thr Thr Leu Cys Leu Val Arg Lys Phe  
 260 265 270  
 Thr Ser Ala Val Arg Glu Glu Leu Phe Lys Ala Leu Gly Leu His Lys  
 275 280 285  
 Leu His Leu Pro Asn Thr Ser Arg Asp Ser Glu Thr Ala Lys Pro Ser  
 290 295 300  
 Val Asn Gly His Gln Lys Ala Leu Xaa  
 305 310

<210> 133  
 <211> 183  
 <212> PRT  
 <213> Homo sapiens

<220>

<221> SITE  
 <222> (183)  
 <223> Xaa equals stop translation

<400> 133

Met Met Val Cys Ser Ile Met Met Tyr Phe Leu Leu Gly Ile Thr Leu  
 1 5 10 15

Leu Arg Ser Tyr Met Gln Ser Val Trp Thr Glu Glu Ser Gln Cys Thr  
 20 25 30

Leu Leu Asn Ala Ser Ile Thr Glu Thr Phe Asn Cys Ser Phe Ser Cys  
 35 40 45

Gly Pro Asp Cys Trp Lys Leu Ser Gln Tyr Pro Cys Leu Gln Val Tyr  
 50 55 60

Val Asn Leu Thr Ser Ser Gly Glu Lys Leu Leu Leu Tyr His Thr Glu  
 65 70 75 80

Glu Thr Ile Lys Ile Asn Gln Lys Cys Ser Tyr Ile Pro Lys Cys Gly  
 85 90 95

Lys Asn Phe Glu Glu Ser Met Ser Leu Val Asn Val Val Met Glu Asn  
 100 105 110

Phe Arg Lys Tyr Gln His Phe Ser Cys Tyr Ser Asp Pro Glu Gly Asn  
 115 120 125

Gln Lys Ser Val Ile Leu Thr Lys Leu Tyr Ser Ser Asn Val Leu Phe  
 130 135 140

His Ser Leu Phe Trp Pro Thr Cys Met Met Ala Gly Gly Val Ala Ile  
 145 150 155 160

Val Ala Met Val Lys Leu Thr Gln Tyr Leu Ser Leu Leu Cys Glu Arg  
 165 170 175

Ile Gln Arg Ile Asn Arg Xaa  
 180

<210> 134  
 <211> 147  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (147)  
 <223> Xaa equals stop translation

<400> 134

Met Trp Lys Leu Trp Arg Ala Glu Glu Gly Ala Ala Ala Leu Gly Gly  
 1 5 10 15

Ala Leu Phe Leu Leu Leu Phe Ala Leu Gly Val Arg Gln Leu Leu Lys  
 20 25 30

Gln Arg Arg Pro Met Gly Phe Pro Pro Gly Pro Pro Gly Leu Pro Phe  
 35 40 45  
 Ile Gly Asn Ile Tyr Ser Leu Ala Ala Ser Ser Glu Leu Pro His Val  
 50 55 60  
 Tyr Met Arg Lys Gln Ser Gln Val Tyr Gly Glu Val Gln Pro Arg Arg  
 65 70 75 80  
 Ala Pro Gly Arg Glu Gly Arg Gln Ala Gly Pro Gly Trp Pro Gly Pro  
 85 90 95  
 Ser Trp Leu Asp Leu Trp Pro Pro Leu Gly Arg Leu Val Gly Thr Ser  
 100 105 110  
 Pro Cys Ala Gly Cys Pro Leu Arg Asp Thr Arg Phe Pro Gly Leu Glu  
 115 120 125  
 Gly Arg Ser Pro Arg Arg Arg Ala Pro Leu Gln Gly Glu Pro Arg Pro  
 130 135 140  
 Cys Arg Xaa  
 145

<210> 135  
 <211> 122  
 <212> PRT  
 <213> Homo sapiens

<400> 135  
 Met Arg Val Arg Ile Gly Leu Thr Leu Leu Leu Cys Ala Val Leu Leu  
 1 5 10 15  
 Ser Leu Ala Ser Ala Ser Ser Asp Glu Glu Gly Ser Gln Asp Glu Ser  
 20 25 30  
 Leu Asp Ser Lys Thr Thr Leu Thr Ser Asp Glu Ser Val Lys Asp His  
 35 40 45  
 Thr Thr Ala Gly Arg Val Val Ala Gly Gln Ile Phe Leu Asp Ser Glu  
 50 55 60  
 Glu Ser Glu Leu Glu Ser Ser Ile Gln Glu Glu Glu Asp Ser Leu Lys  
 65 70 75 80  
 Ser Gln Glu Gly Glu Ser Val Thr Glu Asp Ile Ser Phe Leu Glu Ser  
 85 90 95  
 Pro Asn Pro Glu Asn Lys Asp Tyr Glu Glu Pro Lys Lys Val Arg Lys  
 100 105 110  
 Pro Gly Ser Leu Asp Ile Phe Leu Ala Phe  
 115 120

<210> 136  
 <211> 112  
 <212> PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 136

Met Ala Arg Gly Ser Leu Arg Arg Leu Leu Arg Leu Leu Val Leu Gly  
 1 5 10 15

Leu Trp Leu Ala Leu Leu Arg Ser Val Ala Gly Glu Gln Ala Pro Gly  
 20 25 30

Thr Ala Pro Cys Ser Arg Gly Ser Ser Trp Ser Ala Asp Leu Asp Lys  
 35 40 45

Cys Met Asp Cys Ser Thr Ser Cys Pro Leu Pro Ala Ala Leu Ala His  
 50 55 60

Pro Trp Gly Arg Ser Glu Pro Asp Leu Arg Ala Gly Ala Ala Phe Trp  
 65 70 75 80

Leu Phe Gly Leu Glu Thr Met Pro Gln Arg Glu Lys Phe Thr Thr Pro  
 85 90 95

Ile Glu Glu Thr Gly Gly Glu Gly Cys Pro Ala Val Ala Leu Ile Gln  
 100 105 110

&lt;210&gt; 137

&lt;211&gt; 140

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (140)

&lt;223&gt; Xaa equals stop translation

&lt;400&gt; 137

Met Leu Leu Gly Pro Val Pro Ile Leu His Ile Lys Ser Gln Leu Trp  
 1 5 10 15

Leu Leu Val Leu Ile Leu Val Val Ser Gly Leu Ser Ala Gly Met Ser  
 20 25 30

Ile Ile Pro Thr Phe Pro Glu Ile Leu Ser Cys Ala His Glu Asn Gly  
 35 40 45

Phe Glu Glu Gly Leu Ser Thr Leu Gly Leu Val Ser Gly Leu Phe Ser  
 50 55 60

Ala Met Trp Ser Ile Gly Ala Phe Met Gly Pro Thr Leu Gly Gly Phe  
 65 70 75 80

Leu Tyr Glu Lys Ile Gly Phe Glu Trp Ala Ala Ala Ile Gln Gly Leu  
 85 90 95

Trp Ala Leu Ile Ser Gly Leu Ala Met Gly Leu Phe Tyr Leu Leu Glu  
 100 105 110

Tyr Ser Arg Arg Lys Arg Ser Lys Ser Gln Asn Ile Leu Ser Thr Glu  
 115 120 125

Glu Glu Arg Thr Thr Leu Leu Pro Asn Glu Thr Xaa  
 130 135 140

<210> 138

<211> 404

<212> PRT

<213> Homo sapiens

<400> 138

Met Arg Leu Gln Asp Val Tyr Met Leu Asn Val Lys Gly Leu Ala Arg  
 1 5 10 15

Gly Val Phe Gln Arg Val Thr Gly Ser Ala Ile Thr Asp Leu Tyr Ser  
 20 25 30

Pro Lys Arg Leu Phe Ser Leu Thr Gly Asp Asp Cys Phe Gln Val Gly  
 35 40 45

Lys Val Ala Tyr Asp Met Gly Asp Tyr Tyr His Ala Ile Pro Trp Leu  
 50 55 60

Glu Glu Ala Val Ser Leu Phe Arg Gly Ser Tyr Gly Glu Trp Lys Thr  
 65 70 75 80

Glu Asp Glu Ala Ser Leu Glu Asp Ala Leu Asp His Leu Ala Phe Ala  
 85 90 95

Tyr Phe Arg Ala Gly Asn Val Ser Cys Ala Leu Ser Leu Ser Arg Glu  
 100 105 110

Phe Leu Leu Tyr Ser Pro Asp Asn Lys Arg Met Ala Arg Asn Val Leu  
 115 120 125

Lys Tyr Glu Arg Leu Leu Ala Glu Ser Pro Asn His Val Val Ala Glu  
 130 135 140

Ala Val Ile Gln Arg Pro Asn Ile Pro His Leu Gln Thr Arg Asp Thr  
 145 150 155 160

Tyr Glu Gly Leu Cys Gln Thr Leu Gly Ser Gln Pro Thr Leu Tyr Gln  
 165 170 175

Ile Pro Ser Leu Tyr Cys Ser Tyr Glu Thr Asn Ser Asn Ala Tyr Leu  
 180 185 190

Leu Leu Gln Pro Ile Arg Lys Glu Val Ile His Leu Glu Pro Tyr Ile  
 195 200 205

Ala Leu Tyr His Asp Phe Val Ser Asp Ser Glu Ala Gln Lys Ile Arg  
 210 215 220

Glu Leu Ala Glu Pro Trp Leu Gln Arg Ser Val Val Ala Ser Gly Glu  
 225 230 235 240



<210>	139
<211>	96
<212>	PRT
<213>	Homo sapiens

```

<400> 139
Met Lys Ala Pro His Thr Gly Val Leu His Leu Gly Ser Val Trp Val
  1              5              10              15
Phe Leu Gly Pro Phe Leu Leu Gly Val Gly Tyr Thr Leu Thr Phe Asn
      20              25              30
Pro Leu Ser Gly Cys Met Ser Thr Val Arg Trp Leu Asn Ser Asn Ile
      35              40              45
Thr Ala Asn Arg Thr Leu Ser Arg Ser Val Cys His Val Thr Pro Leu
      50              55              60
His Arg Ser Leu Ser Pro His Asp Gly Glu Tyr Leu Arg Gln Met Leu
      65              70              75              80

```

100

Leu Asn Ser Ser Ser Arg Ala Gly Glu Ala Gly Ser Trp Gly Tyr Xaa  
                     85                    90                    95

<210> 140  
 <211> 240  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (240)  
 <223> Xaa equals stop translation

<400> 140  
 Met Gly Ser Cys Ala Arg Leu Leu Leu Leu Trp Gly Cys Thr Val Val  
   1                    5                    10                    15  
 Ala Ala Gly Leu Ser Gly Val Ala Gly Val Ser Ser Arg Cys Glu Lys  
                     20                    25                    30  
 Ala Cys Asn Pro Arg Met Gly Asn Leu Ala Leu Gly Arg Lys Leu Trp  
                     35                    40                    45  
 Ala Asp Thr Thr Cys Gly Gln Asn Ala Thr Glu Leu Tyr Cys Phe Tyr  
                     50                    55                    60  
 Ser Glu Asn Thr Asp Leu Thr Cys Arg Gln Pro Lys Cys Asp Lys Cys  
   65                    70                    75                    80  
 Asn Ala Ala Tyr Pro His Leu Ala His Leu Pro Ser Ala Met Ala Asp  
                     85                    90                    95  
 Ser Ser Phe Arg Phe Pro Arg Thr Trp Trp Gln Ser Ala Glu Asp Val  
                     100                    105                    110  
 His Arg Glu Lys Ile Gln Leu Asp Leu Glu Ala Glu Phe Tyr Phe Thr  
                     115                    120                    125  
 His Leu Ile Val Met Phe Lys Ser Pro Arg Pro Ala Ala Met Val Leu  
                     130                    135                    140  
 Asp Arg Ser Gln Asp Phe Gly Lys Thr Trp Lys Pro Tyr Lys Tyr Phe  
  145                    150                    155                    160  
 Ala Thr Asn Cys Ser Ala Thr Phe Gly Leu Glu Asp Asp Val Val Lys  
                     165                    170                    175  
 Lys Gly Ala Ile Cys Thr Ser Lys Tyr Ser Ser Pro Phe Pro Cys Thr  
                     180                    185                    190  
 Gly Arg Lys Val Ile Phe Lys Ala Leu Ser Pro Pro Tyr Asp Thr Glu  
                     195                    200                    205  
 Asn Pro Tyr Ser Ala Lys Val Gln Glu Gln Leu Lys Ile Thr Asn Leu

210

215

220

Pro Arg Ala Ala Ala Glu Thr Thr Val Leu Ser Leu Ser Glu Lys Xaa  
 225 230 235 240

&lt;210&gt; 141

&lt;211&gt; 54

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (54)

&lt;223&gt; Xaa equals stop translation

&lt;400&gt; 141

Met Met Ile Ser Gly Leu Lys Leu Leu Val Leu Phe Leu Lys Phe Ala  
 1 5 10 15

Pro Glu Asn Tyr Cys Leu Ser Thr Glu Thr Leu Gln Met Pro Asn Arg  
 20 25 30

His Leu Arg Leu Ser Lys Ala Thr Cys Tyr Leu Met Lys Cys Leu Leu  
 35 40 45

Pro Ser Tyr Phe Glu Xaa  
 50

&lt;210&gt; 142

&lt;211&gt; 67

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (67)

&lt;223&gt; Xaa equals stop translation

&lt;400&gt; 142

Met Arg Ser Leu Ile Ser Ser His Pro Cys Gln His Leu Leu Leu Leu  
 1 5 10 15

Leu Leu Leu Leu Phe Leu Ile Leu Ala Ile Leu Val Asp Val Lys Trp  
 20 25 30

Tyr Leu Val Leu Phe Ile Cys Ile Ser Leu Met Thr Ser Asp Val Glu  
 35 40 45

His Leu Phe Met Cys Leu Leu Ala Ile Arg Ile Ser Ser Trp Arg Asn  
 50 55 60

Val Tyr Xaa  
 65

<210> 143  
 <211> 108  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (48)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (55)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (58)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (67)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<400> 143  
 Met Phe Tyr Lys Leu Thr Leu Ile Leu Cys Glu Leu Ser Val Ala Gly  
     1                    5                    10                    15  
 Val Thr Gln Ala Ala Ser Gln Arg Pro Leu Gln Arg Leu Pro Arg His  
                     20                    25                    30  
 Ile Cys Ser Gln Arg Asn Pro Pro Gly Arg Cys Leu Leu Lys Ala Xaa  
                     35                    40                    45  
 Leu Gln Thr Thr Trp Gly Xaa Pro Asp Xaa Gln Phe Pro Gly Cys Pro  
                     50                    55                    60  
 His Pro Xaa Arg Val Thr Leu Asn Ala Arg Gln Met Gly Asn Gly Lys  
                     65                    70                    75                    80  
 Glu Lys Lys Ala Ala Asp Leu Lys Leu Lys Phe Pro Gln Lys Arg Phe  
                     85                    90                    95  
 Tyr Leu Ser Ala Phe Ser Glu Arg Ile Lys Ala Phe  
                     100                    105

<210> 144  
 <211> 84  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (84)  
 <223> Xaa equals stop translation

103

&lt;400&gt; 144

Met Ala Ser Val Gly Thr Thr Leu Val Ser Pro Leu Leu Cys Leu Leu  
 1 5 10 15

Ile Pro Thr Arg Val Ser Asp Pro Trp Leu Gln Asn Thr Pro Leu His  
 20 25 30

Pro Trp Lys Thr Ile Thr Ile Ile Asp Tyr Tyr Leu Ser Leu Gly Phe  
 35 40 45

Leu Gly Trp Thr Gly Leu Ser Trp Val Val His Phe Gly Ala Ser Ala  
 50 55 60

Val Met Gly Arg Gln Trp Leu Gly Ser Leu Gln Arg Leu Pro Cys Ile  
 65 70 75 80

Ser Gly Ser Xaa

&lt;210&gt; 145

&lt;211&gt; 166

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 145

Met Gly Ser Arg Phe Leu Leu Val Leu Leu Ser Gly Leu Thr Val Leu  
 1 5 10 15

Leu Ala Leu Pro Gly Ser Glu Ala Lys Asn Ser Gly Ala Ser Cys Pro  
 20 25 30

Pro Cys Pro Lys Tyr Ala Ser Cys His Asn Ser Thr His Cys Thr Cys  
 35 40 45

~~Glu Asp Gly Phe Arg Ala Arg Ser Gly Arg Thr Tyr Phe His Asp Ser~~  
 50 55 60

Ser Glu Lys Cys Glu Asp Ile Asn Glu Cys Glu Thr Gly Leu Ala Lys  
 65 70 75 80

Cys Lys Tyr Lys Ala Tyr Cys Arg Asn Lys Val Gly Gly Tyr Ile Cys  
 85 90 95

Ser Cys Leu Val Lys Tyr Thr Leu Phe Asn Phe Leu Ala Gly Ile Ile  
 100 105 110

Asp Tyr Asp His Pro Asp Cys Tyr Glu Asn Asn Ser Gln Gly Thr Thr  
 115 120 125

Gln Ser Asn Val Asp Ile Trp Val Ser Gly Val Lys Pro Gly Phe Gly  
 130 135 140

Lys Gln Leu Val Arg Ile Thr Met Pro Phe Ser Tyr Pro Asn Ile Asn  
 145 150 155 160

Met Ser Ser Cys Asp Phe  
 165

<210> 146  
 <211> 70  
 <212> PRT  
 <213> Homo sapiens

<400> 146  
 Met Lys Pro Lys His Leu Glu Trp Cys Leu Ala His Ser Trp Cys Val  
     1                    5                    10                    15  
 Ile Trp Leu Ser Phe Val Ser Pro Pro Thr Ser His Leu Glu Cys Asp  
             20                    25                    30  
 Gly Phe Pro Gly Ser Leu Leu Pro Pro Cys Glu Glu Gly Arg Cys Phe  
             35                    40                    45  
 Pro Phe Thr Phe His His His Asp Cys His Gly Cys Ser Pro Leu Gln  
             50                    55                    60  
 Ser Ser Pro Gly Gln His  
     65                    70

<210> 147  
 <211> 412  
 <212> PRT  
 <213> Homo sapiens

<400> 147  
 Met Cys Cys Trp Pro Leu Leu Leu Leu Trp Gly Leu Leu Pro Gly Thr  
     1                    5                    10                    15  
 Ala Ala Gly Gly Ser Gly Arg Thr Tyr Pro His Arg Thr Leu Leu Asp  
             20                    25                    30  
 Ser Glu Gly Lys Tyr Trp Leu Gly Trp Ser Gln Arg Gly Ser Gln Ile  
             35                    40                    45  
 Ala Phe Arg Leu Gln Val Arg Thr Ala Gly Tyr Val Gly Phe Gly Phe  
             50                    55                    60  
 Ser Pro Thr Gly Ala Met Ala Ser Ala Asp Ile Val Val Gly Gly Val  
     65                    70                    75                    80  
 Ala His Gly Arg Pro Tyr Leu Gln Asp Tyr Phe Thr Asn Ala Asn Arg  
             85                    90                    95  
 Glu Leu Lys Lys Asp Ala Gln Gln Asp Tyr His Leu Glu Tyr Ala Met  
             100                    105                    110  
 Glu Asn Ser Thr His Thr Ile Ile Glu Phe Thr Arg Glu Leu His Thr  
             115                    120                    125  
 Cys Asp Ile Asn Asp Lys Ser Ile Thr Asp Ser Thr Val Arg Val Ile  
             130                    135                    140  
 Trp Ala Tyr His His Glu Asp Ala Gly Glu Ala Gly Pro Lys Tyr His  
     145                    150                    155                    160

Asp Ser Asn Arg Gly Thr Lys Ser Leu Arg Leu Leu Asn Pro Glu Lys  
                             165                            170                            175  
 Thr Ser Val Leu Ser Thr Ala Leu Pro Tyr Phe Asp Leu Val Asn Gln  
                             180                            185                            190  
 Asp Val Pro Ile Pro Asn Lys Asp Thr Thr Tyr Trp Cys Gln Met Phe  
                             195                            200                            205  
 Lys Ile Pro Val Phe Gln Glu Lys His His Val Ile Lys Val Glu Pro  
                             210                            215                            220  
 Val Ile Gln Arg Gly His Glu Ser Leu Val His His Ile Leu Leu Tyr  
                             225                            230                            235                            240  
 Gln Cys Ser Asn Asn Phe Asn Asp Ser Val Leu Glu Ser Gly His Glu  
                             245                            250                            255  
 Cys Tyr His Pro Asn Met Pro Asp Ala Phe Leu Thr Cys Glu Thr Val  
                             260                            265                            270  
 Ile Phe Ala Trp Ala Ile Gly Gly Glu Gly Phe Ser Tyr Pro Pro His  
                             275                            280                            285  
 Val Gly Leu Ser Leu Gly Thr Pro Leu Asp Pro His Tyr Val Leu Leu  
                             290                            295                            300  
 Glu Val His Tyr Asp Asn Pro Thr Tyr Glu Glu Gly Leu Ile Asp Asn  
                             305                            310                            315                            320  
 Ser Gly Leu Arg Leu Phe Tyr Thr Met Asp Ile Arg Lys Tyr Asp Ala  
                             325                            330                            335  
 Gly Val Ile Glu Ala Gly Leu Trp Val Ser Leu Phe His Thr Ile Pro  
                             340                            345                            350

---

Pro Gly Met Pro Glu Phe Gln Ser Glu Gly His Cys Thr Leu Glu Cys  
                             355                            360                            365  
 Leu Glu Glu Leu Trp Lys Pro Lys Ser Gln Val Glu Phe Met Cys Leu  
                             370                            375                            380  
 Leu Phe Phe Ser Met Leu Thr Trp Leu Ala Glu His Gln Ala Ala Ser  
                             385                            390                            395                            400  
 Phe Ser Lys Arg Glu Gly Asn Glu Ile Thr Cys Leu  
                             405                            410

&lt;210&gt; 148

&lt;211&gt; 85

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (85)

&lt;223&gt; Xaa equals stop translation

&lt;400&gt; 148

Met Asn Val Phe Leu Pro Pro Ala Leu Gly Thr Trp Gly Val Ala Arg  
 1 5 10 15

Phe Phe Pro His Leu Val Pro Glu Arg Trp Cys Leu Val Phe Cys Cys  
 20 25 30

Trp Ile Phe Phe Phe Phe Phe Phe Cys Thr Lys Val Ala Thr Arg  
 35 40 45

Ser Val Leu Gly Asp Gln Ala Gly Leu Gly Val Gly Gly Pro His Leu  
 50 55 60

Pro Leu Pro Gly Ser His Ser Val Ser Val Pro Glu Lys Thr Ile Phe  
 65 70 75 80

Ser Leu Lys Gln Xaa  
 85

&lt;210&gt; 149

&lt;211&gt; 154

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (154)

&lt;223&gt; Xaa equals stop translation

&lt;400&gt; 149

Met Gly Arg Leu Pro Leu Leu Arg Arg Val Leu Lys Gly Leu Gln Leu  
 1 5 10 15

Leu Leu Ser Leu Leu Ala Phe Ile Cys Glu Glu Val Val Ser Gln Cys  
 20 25 30

Thr Leu Cys Gly Gly Leu Tyr Phe Phe Glu Phe Val Ser Cys Ser Ala  
 35 40 45

Phe Leu Leu Ser Leu Leu Ile Leu Ile Val Tyr Cys Thr Pro Phe Tyr  
 50 55 60

Glu Arg Val Asp Thr Thr Lys Val Lys Ser Ser Asp Phe Tyr Ile Thr  
 65 70 75 80

Leu Gly Thr Gly Cys Val Phe Leu Leu Ala Ser Ile Ile Phe Val Ser  
 85 90 95

Thr His Asp Arg Thr Ser Ala Glu Ile Ala Ala Ile Val Phe Gly Phe  
 100 105 110

Ile Ala Ser Phe Met Phe Leu Leu Asp Phe Ile Thr Met Leu Tyr Glu  
 115 120 125

Lys Arg Gln Glu Ser Gln Leu Arg Lys Pro Glu Asn Thr Thr Arg Ala  
 130 135 140

Glu Ala Leu Thr Glu Pro Leu Asn Ala Xaa



107

145

150

<210> 150  
 <211> 130  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (130)  
 <223> Xaa equals stop translation

&lt;400&gt; 150

Met Arg Gly His Leu Ala Gly Phe Pro Ala Leu Ser Gly Leu Ala Ser  
 1 5 10 15

Val Cys Leu Trp Ala Thr Phe Ser Ala Gln Leu Pro Gly Pro Val Ala  
 20 25 30

Ala Thr Ser Trp Thr Pro Ala Pro Leu Gly Cys Ser Ala Ala Arg Ser  
 35 40 45

Gly Pro Glu Lys Arg Leu Gly Thr Ala Ala Pro Gly Ser Ala Ala Ser  
 50 55 60

Leu Ala Gln Ala Gly Pro Gly Ala Pro Cys Arg Val Leu Pro Val Asp  
 65 70 75 80

Pro Ala Pro Ala Ala Leu Asn Val Arg Glu Pro Gly Trp Leu Gly Gly  
 85 90 95

Leu Phe Asp Gly Ala Leu Leu Gln Val Leu Leu Asn Phe Leu Arg Lys  
 100 105 110

Ser Thr Asp Val Leu Met Asp Thr Arg Glu Ala Glu Ser Leu Glu Val  
 115 120 125

Glu Xaa  
 130

<210> 151  
 <211> 62  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (62)  
 <223> Xaa equals stop translation

&lt;400&gt; 151

Met Leu Phe Trp Ala Tyr Pro Ile Cys Val Phe Ile Asp Ser Leu Ser  
 1 5 10 15

Cys Gln Pro Cys Leu Trp Ser Thr Gly Ala Thr Ser His Phe Asn Ser  
 20 25 30

108

Pro Thr Thr Ser Pro Leu Phe Thr Leu Phe Met Pro Cys Ala Leu Ala  
                   35                                  40                                  45

Pro Asn Pro Phe Thr Gln Leu Gly Lys Leu Asp Asp Arg Xaa  
           50                                  55                                  60

<210> 152  
 <211> 225  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (225)  
 <223> Xaa equals stop translation

<400> 152  
 Met Gly Ile Phe Pro Gly Ile Ile Leu Ile Phe Leu Arg Val Lys Phe  
   1                                  5                                  10                                  15

Ala Thr Ala Ala Val Ile Val Ser Gly His Gln Lys Ser Thr Thr Val  
                   20                                  25                                  30

Ser His Glu Met Ser Gly Leu Asn Trp Lys Pro Phe Val Tyr Gly Gly  
           35                                  40                                  45

Leu Ala Ser Ile Val Ala Glu Phe Gly Thr Phe Pro Val Asp Leu Thr  
           50                                  55                                  60

Lys Thr Arg Leu Gln Val Gln Gly Gln Ser Ile Asp Ala Arg Phe Lys  
           65                                  70                                  75                                  80

Glu Ile Lys Tyr Arg Gly Met Phe His Ala Leu Phe Arg Ile Cys Lys  
                   85                                  90                                  95

Glu Glu Gly Val Leu Ala Leu Tyr Ser Gly Ile Ala Pro Ala Leu Leu  
                   100                                  105                                  110

Arg Gln Ala Ser Tyr Gly Thr Ile Lys Ile Gly Ile Tyr Gln Ser Leu  
           115                                  120                                  125

Lys Arg Leu Phe Val Glu Arg Leu Glu Asp Glu Thr Leu Leu Ile Asn  
           130                                  135                                  140

Met Ile Cys Gly Val Val Ser Gly Val Ile Ser Ser Thr Ile Ala Asn  
           145                                  150                                  155                                  160

Pro Thr Asp Val Leu Lys Ile Arg Met Gln Ala Gln Gly Ser Leu Phe  
                   165                                  170                                  175

Gln Gly Ser Met Ile Gly Ser Phe Ile Asp Ile Tyr Gln Gln Glu Gly  
                   180                                  185                                  190

Thr Arg Gly Leu Trp Arg Val Ser Thr Leu Phe Leu Leu Ser Tyr  
           195                                  200                                  205

Thr Leu Ser Ser Tyr Asn Leu Gln Arg Ile Phe Phe Tyr Ile Lys Thr  
           210                                  215                                  220

Xaa  
225

<210> 153  
<211> 69  
<212> PRT  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (69)  
<223> Xaa equals stop translation

<400> 153

Met Leu Met Leu Leu Thr Leu Leu Val Leu Gly Met Val Trp Val Ala  
1 5 10 15

Ser Ala Ile Val Asp Lys Asn Lys Ala Asn Arg Glu Ser Leu Tyr Asp  
20 25 30

Phe Trp Glu Tyr Tyr Leu Pro Tyr Leu Tyr Ser Cys Ile Ser Phe Leu  
35 40 45

Gly Val Leu Leu Leu Leu Ala Ala Gly Arg Pro Gly Gly Ala Ala Val  
50 55 60

Leu Leu Ser Leu Xaa  
65

<210> 154  
<211> 84  
<212> PRT  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (84)  
<223> Xaa equals stop translation

<400> 154

Met Tyr Gly Val Cys Leu Cys Val Ile Val Cys Val Ser Gly Val Ser  
1 5 10 15

Leu Cys Leu Tyr Val Trp Gly Val Ser Val Cys Asp Cys Val Ser Val  
20 25 30

Phe Met Cys Val Cys Leu Cys Val Ile Phe Cys Val Tyr Gly Lys Pro  
35 40 45

Arg Thr Glu His Tyr His Ser Pro His Leu Ala Lys Gln Lys Ala Phe  
50 55 60

Arg Glu Met Cys Gly Arg His Asp Val Ser Ala Ala Gly Ile Phe Gln  
65 70 75 80

Ser Tyr Val Xaa

<210> 155  
 <211> 61  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (61)  
 <223> Xaa equals stop translation

<400> 155  
 Met His Val Leu Leu Phe Ser Phe Leu Ile Pro Phe Leu Leu Leu Ser  
     1                    5                    10                    15  
 Pro Val Gly Val Thr Cys Asn Ser His Met Leu Glu Arg Gln Val Ser  
                     20                    25                    30  
 Trp Leu Lys Lys Arg Ser Thr Gln Ala Ser Gln Gln Phe Asn Lys Phe  
                     35                    40                    45  
 Leu Arg Gly Ile Ser Asn Val Gly Arg Ile Val Ile Xaa  
     50                    55                    60

<210> 156  
 <211> 84  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (84)  
 <223> Xaa equals stop translation

<400> 156  
 Met Cys Leu Leu Val Glu Tyr Ser Leu Met Ile Leu Thr Ile Ile Pro  
     1                    5                    10                    15  
 Ser Leu Leu Ser Phe Val Leu Cys Leu Lys Gly Ile Lys His Gly Asn  
                     20                    25                    30  
 Tyr Ile Phe Gln Thr Pro Leu Pro Glu Gly Tyr Gly Trp Ile Ser Ala  
                     35                    40                    45  
 Met Ser Gly Leu Cys Ile Lys Phe Gly Arg Arg Lys Arg Arg Lys Thr  
     50                    55                    60  
 Trp Leu Leu Gln Val Gly Thr Leu Ala Thr Ile Asp Thr Glu Phe Ala  
     65                    70                    75                    80  
 Arg Ser Cys Xaa

<210> 157  
 <211> 162

111

<212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (162)  
 <223> Xaa equals stop translation

<400> 157

Met Ala Leu Ser Leu Thr Leu Cys Phe Val Met Phe Trp Thr Pro Asn  
 1 5 10 15

Val Ser Glu Lys Ile Leu Ile Asp Ile Ile Gly Val Asp Phe Ala Phe  
 20 25 30

Ala Glu Leu Cys Val Val Pro Leu Arg Ile Phe Ser Phe Phe Pro Val  
 35 40 45

Pro Val Thr Val Arg Ala His Leu Thr Gly Trp Leu Met Thr Leu Lys  
 50 55 60

Lys Thr Phe Val Leu Ala Pro Ser Ser Val Leu Arg Ile Ile Val Leu  
 65 70 75 80

Ile Ala Ser Leu Val Val Leu Pro Tyr Leu Gly Val His Gly Ala Thr  
 85 90 95

Leu Gly Val Gly Ser Leu Leu Ala Gly Phe Val Gly Glu Ser Thr Met  
 100 105 110

Val Ala Ile Ala Ala Cys Tyr Val Tyr Arg Lys Gln Lys Lys Lys Met  
 115 120 125

Glu Asn Glu Ser Ala Thr Glu Gly Glu Asp Ser Ala Met Thr Asp Met  
 130 135 140

Pro Pro Thr Glu Glu Val Thr Asp Ile Val Glu Met Arg Glu Glu Asn  
 145 150 155 160

Glu Xaa

<210> 158  
 <211> 146  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (96)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (107)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE  
 <222> (111)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (115)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (122)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (132)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<400> 158  
 Met Glu Pro Gln Leu Gly Pro Glu Ala Ala Ala Leu Arg Pro Gly Trp  
   1                  5                  10                  15  
 Leu Ala Leu Leu Leu Trp Val Ser Ala Leu Ser Cys Ser Phe Ser Leu  
                   20                  25                  30  
 Pro Ala Ser Ser Leu Ser Ser Leu Val Pro Gln Val Arg Thr Ser Tyr  
           35                  40                  45  
 Asn Phe Gly Arg Thr Phe Leu Gly Leu Asp Lys Cys Asn Ala Cys Ile  
   50                  55                  60  
 Gly Thr Ser Ile Cys Lys Lys Phe Phe Lys Glu Glu Ile Arg Ser Asp  
   65                  70                  75                  80  
 Asn Trp Leu Ala Ser His Leu Gly Thr Ala Ser Arg Phe Pro Leu Xaa  
                   85                  90                  95  
 Ser Tyr Pro Cys Lys Leu Leu Gln Met Ile Xaa Lys Ile Trp Xaa Pro  
           100                  105                  110  
 Cys Gly Xaa Leu Leu Thr Gly Gln Gln Xaa Ser Asn Glu Ile Ser Lys  
   115                  120                  125  
 Gln Glu Ile Xaa Cys Leu Leu His Pro Pro Pro Lys Asn Leu His Ile  
   130                  135                  140  
 Asp Val  
 145

<210> 159  
 <211> 143  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (143)

113

&lt;223&gt; Xaa equals stop translation

&lt;400&gt; 159

Met Trp Trp Ala Val Met Gly Gly Val Ile Gly Ser Trp Leu Ser Pro  
 1 5 10 15

Leu Ser Ile Ala Glu Cys Cys His Asp Leu Trp Thr Ser Gln Ser Cys  
 20 25 30

Glu His Ala Gly Ala Leu Cys Gly Asp Leu Leu Cys Ala Cys Arg Lys  
 35 40 45

Val Gly Val Trp Cys Ala Leu Gln Gln His Trp Trp Asn Arg Cys Val  
 50 55 60

Cys Pro His Ala Val Ile Arg Val His Cys Thr Gly Ala Ser Tyr Thr  
 65 70 75 80

Leu Gln Lys Ile Cys Ser Cys Asn Pro Lys Phe Met Gly Arg His Pro  
 85 90 95

His Arg Trp Gln Gln Ile Arg Lys Cys Ser Gln Pro Val Leu Arg Gly  
 100 105 110

Ser Arg Ala Ala Phe Ile Trp Val Arg Leu Ala Ala Leu Asn Phe Ile  
 115 120 125

Ser Ser Phe Arg Cys Ile Ser Leu Ile Ser Tyr Ser Ala Phe Xaa  
 130 135 140

&lt;210&gt; 160

&lt;211&gt; 51

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (51)

&lt;223&gt; Xaa equals stop translation

&lt;400&gt; 160

Met Lys Val Ser Asp Phe Asn Phe Leu Ile Phe Leu Ile Phe Ala Leu  
 1 5 10 15

Phe Leu Thr Leu Glu Ala Phe Leu Lys Phe Thr Lys Arg Val Leu Ala  
 20 25 30

Val Val Gly Asn Leu Pro Glu Pro Pro Ile Ile Lys Thr Ile Gly Phe  
 35 40 45

Leu Tyr Xaa  
 50

&lt;210&gt; 161

&lt;211&gt; 65

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

<220>  
 <221> SITE  
 <222> (65)  
 <223> Xaa equals stop translation

<400> 161  
 Met Val Trp Ser Ala Ala Pro Ala Pro Cys Cys Leu Leu Gly Val Leu  
     1                    5                    10                    15  
 Gly Leu Val Gln Val Leu Gly Ala Gln Ala Val Gly Pro Trp Thr Ala  
             20                    25                    30  
 Ser Ala Cys Leu Gly Ala Ala Gln Ala Gln Pro Cys Arg Pro Cys Lys  
             35                    40                    45  
 Glu Ser Ser Leu Arg Leu Phe Ser Ala Ser Ala Pro Ser Met Thr His  
     50                    55                    60  
 Xaa  
   65

<210> 162  
 <211> 59  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (59)  
 <223> Xaa equals stop translation

<400> 162  
 Met Glu Lys Tyr Cys Leu Gly Asn Asn Met Leu Ser Arg Phe Cys Leu  
     1                    5                    10                    15  
 Phe Leu Ile Met Leu Leu His Ile Leu Leu Phe Leu Val Ile Phe Ile  
             20                    25                    30  
 Gln Arg His Thr Val Val Ser Leu Ser Lys His His Pro Phe Val Pro  
             35                    40                    45  
 Thr Asn Gly Ser Lys Ser Tyr Ser Ser Phe Xaa  
     50                    55

<210> 163  
 <211> 374  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (84)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE



115

&lt;222&gt; (112)

&lt;223&gt; Xaa equals any of the naturally occurring L-amino acids

&lt;400&gt; 163

Met Arg Pro Gly Thr Ala Leu Gln Ala Val Leu Leu Ala Val Leu Leu  
 1 5 10 15

Val Gly Leu Arg Ala Ala Thr Gly Arg Leu Leu Ser Gly Gln Pro Val  
 20 25 30

Cys Arg Gly Gly Thr Gln Arg Pro Cys Tyr Lys Val Ile Tyr Phe His  
 35 40 45

Asp Thr Ser Arg Arg Leu Asn Phe Glu Glu Ala Lys Glu Ala Cys Arg  
 50 55 60

Arg Asp Gly Gly Gln Leu Val Ser Ile Glu Ser Glu Asp Glu Gln Lys  
 65 70 75 80

Leu Ile Glu Xaa Phe Ile Glu Asn Leu Leu Pro Ser Asp Gly Asp Phe  
 85 90 95

Trp Ile Gly Leu Arg Arg Arg Glu Glu Lys Gln Ser Asn Ser Thr Xaa  
 100 105 110

Cys Gln Asp Leu Tyr Ala Trp Thr Asp Gly Ser Ile Ser Gln Phe Arg  
 115 120 125

Asn Trp Tyr Val Asp Glu Pro Ser Cys Gly Ser Glu Val Cys Val Val  
 130 135 140

Met Tyr His Gln Pro Ser Ala Pro Ala Gly Ile Gly Gly Pro Tyr Met  
 145 150 155 160

Phe Gln Trp Asn Asp Asp Arg Cys Asn Met Lys Asn Asn Phe Ile Cys  
 165 170 175

Lys Tyr Ser Asp Glu Lys Pro Ala Val Pro Ser Arg Glu Ala Glu Gly  
 180 185 190

Glu Glu Thr Glu Leu Thr Thr Pro Val Leu Pro Glu Glu Thr Gln Glu  
 195 200 205

Glu Asp Ala Lys Lys Thr Phe Lys Glu Ser Arg Glu Ala Ala Leu Asn  
 210 215 220

Leu Ala Tyr Ile Leu Ile Pro Ser Ile Pro Leu Leu Leu Leu Val  
 225 230 235 240

Val Thr Thr Val Val Cys Trp Val Trp Ile Cys Arg Lys Arg Lys Arg  
 245 250 255

Glu Gln Pro Asp Pro Ser Thr Lys Lys Gln His Thr Ile Trp Pro Ser  
 260 265 270

Pro His Gln Gly Asn Ser Pro Asp Leu Glu Val Tyr Asn Val Ile Arg  
 275 280 285

Lys Gln Ser Glu Ala Asp Leu Ala Glu Thr Arg Pro Asp Leu Lys Asn

116

290                      295                      300  
 Ile Ser Phe Arg Val Cys Ser Gly Glu Ala Thr Pro Asp Asp Met Ser  
 305                      310                      315                      320  
 Cys Asp Tyr Asp Asn Met Ala Val Asn Pro Ser Glu Ser Gly Phe Val  
                     325                      330                      335  
 Thr Leu Val Ser Val Glu Ser Gly Phe Val Thr Asn Asp Ile Tyr Glu  
                     340                      345                      350  
 Phe Ser Pro Asp Gln Met Gly Arg Ser Lys Glu Ser Gly Trp Val Glu  
                     355                      360                      365  
 Asn Glu Ile Tyr Gly Tyr  
 370

<210> 164  
 <211> 64  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (64)  
 <223> Xaa equals stop translation

<400> 164  
 Met His Pro Gln Leu Ile Pro Ser Val Ile Ala Val Val Phe Ile Leu  
 1                      5                      10                      15  
 Leu Leu Gly Val Cys Phe Ile Ala Ser Cys Leu Val Thr His His Asn  
                     20                      25                      30  
 Phe Ser Arg Cys Lys Arg Gly Thr Gly Val His Lys Leu Glu His His  
                     35                      40                      45  
 Ala Lys Leu Lys Cys Ile Lys Glu Lys Ser Glu Leu Lys Ser Cys Xaa  
                     50                      55                      60

<210> 165  
 <211> 743  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (743)  
 <223> Xaa equals stop translation

<400> 165  
 Met Ala Val Arg Glu Leu Cys Phe Pro Arg Gln Arg Gln Val Leu Phe  
 1                      5                      10                      15

Leu Phe Leu Phe Trp Gly Val Ser Leu Ala Gly Ser Gly Phe Gly Arg  
                   20                                  25                                  30  
 Tyr Ser Val Thr Glu Glu Thr Glu Lys Gly Ser Phe Val Val Asn Leu  
                   35                                  40                                  45  
 Ala Lys Asp Leu Gly Leu Ala Glu Gly Glu Leu Ala Ala Arg Gly Thr  
                   50                                  55                                  60  
 Arg Val Val Ser Asp Asp Asn Lys Gln Tyr Leu Leu Leu Asp Ser His  
                   65                                  70                                  75                                  80  
 Thr Gly Asn Leu Leu Thr Asn Glu Lys Leu Asp Arg Glu Lys Leu Cys  
                                   85                                  90                                  95  
 Gly Pro Lys Glu Pro Cys Met Leu Tyr Phe Gln Ile Leu Met Asp Asp  
                                   100                                  105                                  110  
 Pro Phe Gln Ile Tyr Arg Ala Glu Leu Arg Val Arg Asp Ile Asn Asp  
                                   115                                  120                                  125  
 His Ala Pro Val Phe Gln Asp Lys Glu Thr Val Leu Lys Ile Ser Glu  
                   130                                  135                                  140  
 Asn Thr Ala Glu Gly Thr Ala Phe Arg Leu Glu Arg Ala Gln Asp Pro  
                   145                                  150                                  155                                  160  
 Asp Gly Gly Leu Asn Gly Ile Gln Asn Tyr Thr Ile Ser Pro Asn Ser  
                                   165                                  170                                  175  
 Phe Phe His Ile Asn Ile Ser Gly Gly Asp Glu Gly Met Ile Tyr Pro  
                   180                                  185                                  190  
 Glu Leu Val Leu Asp Lys Ala Leu Asp Arg Glu Glu Gln Gly Glu Leu  
                   195                                  200                                  205

---

Ser Leu Thr Leu Thr Ala Leu Asp Gly Gly Ser Pro Ser Arg Ser Gly  
                   210                                  215                                  220  
 Thr Ser Thr Val Arg Ile Val Val Leu Asp Val Asn Asp Asn Ala Pro  
                   225                                  230                                  235                                  240  
 Gln Phe Ala Gln Ala Leu Tyr Glu Thr Gln Ala Pro Glu Asn Ser Pro  
                                   245                                  250                                  255  
 Ile Gly Phe Leu Ile Val Lys Val Trp Ala Glu Asp Val Asp Ser Gly  
                   260                                  265                                  270  
 Val Asn Ala Glu Val Ser Tyr Ser Phe Phe Asp Ala Ser Glu Asn Ile  
                   275                                  280                                  285  
 Arg Thr Thr Phe Gln Ile Asn Pro Phe Ser Gly Glu Ile Phe Leu Arg  
                   290                                  295                                  300  
 Glu Leu Leu Asp Tyr Glu Leu Val Asn Ser Tyr Lys Ile Asn Ile Gln  
                   305                                  310                                  315                                  320  
 Ala Met Asp Gly Gly Gly Leu Ser Ala Arg Cys Arg Val Leu Val Glu  
                                   325                                  330                                  335

Val Leu Asp Thr Asn Asp Asn Pro Pro Glu Leu Ile Val Ser Ser Phe  
 340 345 350  
 Ser Asn Ser Val Ala Glu Asn Ser Pro Glu Thr Pro Leu Ala Val Phe  
 355 360 365  
 Lys Ile Asn Asp Arg Asp Ser Gly Glu Asn Gly Lys Met Val Cys Tyr  
 370 375 380  
 Ile Gln Glu Asn Leu Pro Phe Leu Leu Lys Pro Ser Val Glu Asn Phe  
 385 390 395 400  
 Tyr Ile Leu Ile Thr Glu Gly Ala Leu Asp Arg Glu Ile Arg Ala Glu  
 405 410 415  
 Tyr Asn Ile Thr Ile Thr Val Thr Asp Leu Gly Thr Pro Arg Leu Lys  
 420 425 430  
 Thr Glu His Asn Ile Thr Val Leu Val Ser Asp Val Asn Asn Ala  
 435 440 445  
 Pro Ala Phe Thr Gln Thr Ser Tyr Thr Leu Phe Val Arg Glu Asn Asn  
 450 455 460  
 Ser Pro Ala Leu His Ile Gly Ser Val Ser Ala Thr Asp Arg Asp Ser  
 465 470 475 480  
 Gly Thr Asn Ala Gln Val Thr Tyr Ser Leu Leu Pro Pro Gln Asp Pro  
 485 490 495  
 His Leu Pro Leu Ala Ser Leu Val Ser Ile Asn Ala Asp Asn Gly His  
 500 505 510  
 Leu Phe Ala Leu Arg Ser Leu Asp Tyr Glu Ala Leu Gln Ala Phe Glu  
 515 520 525  
 Phe Arg Val Gly Ala Thr Asp Arg Gly Ser Pro Ala Leu Asn Ser Glu  
 530 535 540  
 Ala Leu Gly Ala Arg Ala Gly Ala Gly Arg Gln Arg Gln Leu Ala Leu  
 545 550 555 560  
 Arg Ala Val Pro Ala Ala Glu Arg Leu Arg Ala Leu His Arg Ala Gly  
 565 570 575  
 Ala Pro Gly Gly Arg Ala Gly Leu Pro Gly Asp Gln Gly Gly Gly Gly  
 580 585 590  
 Gly Arg Arg Leu Gly Pro Glu Arg Leu Ala Val Val Pro Ala Ala Gln  
 595 600 605  
 Gly His Gly Ala Arg Ala Val Arg Cys Val Gly Ala Gln Trp Gly Gly  
 610 615 620  
 Ala His Arg Gln Ala Ala Glu Arg Ala Arg Arg Ser Gln Ala Gln Ala  
 625 630 635 640  
 Gly Gly Ala Cys Gln Gly Gln Trp Arg Ala Ser Ser Leu Gly His Arg

Gly Gly Arg Glu Gly Arg Xaa  
740

Ser Arg Ile His Arg Glu Glu Asp Phe Gln Phe Ile Leu Lys Gly Ile  
130 135 140

Ala Arg Leu Leu Ser Asn Pro Leu Leu Gln Thr Tyr Leu Pro Asn Ser  
145 150 155 160

Thr Lys Lys Asp Pro Val Pro Pro Gly Ala Ala Ser Ser Leu Leu Glu  
165 170 175

Ala Leu Arg Leu Gln Gln Glu Ile Pro Leu Leu Arg Ala Glu Glu Gln  
180 185 190

Arg Arg Pro Arg His Pro Cys Pro His Pro Leu Leu Pro Gln Arg Cys  
195 200 205

Pro Gly Arg Ser Val Xaa  
210

<210> 167

<211> 213

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (213)

<223> Xaa equals stop translation

<400> 167

Met Pro Ser Leu Arg Phe Leu Ala Leu Ala Leu Leu Leu Ala Ile Leu  
1 5 10 15

Pro Ala Leu Pro Asn Ala His Ala Ala Pro Gly Ile Gly Gly Leu Ile  
20 25 30

Gly Gly Gly Ser Gln Ala Ser Ala Lys Glu Glu Pro Gln Ser Asn Ala  
35 40 45

Gln Pro Ser Ala Asp Glu Arg Lys Gln Arg Leu Leu Ser Gln Ala Glu  
50 55 60

Glu Thr Arg Gln Arg Leu Thr Asp Leu Lys Ala Glu Leu Ala Gly Ala  
65 70 75 80

Pro Lys Glu Ile Ser Glu Ala Gln Arg Thr Leu Ser Lys Leu Val Ser  
85 90 95

Glu Asp Asn Ser Asp Leu Pro Glu Arg Leu Ser Lys Leu Ser Val Pro  
100 105 110

Val Leu Glu Gln Arg Leu Ala Ala Arg Val Asp Glu Leu Ala Leu Trp  
115 120 125

Gln Gln Ala Leu Ser Ala Ala Asn Ser Met Leu Ile Ser Ala Gln Thr  
130 135 140

Arg Pro Glu Arg Ala Gln Ala Asp Ile Ser Lys Asn Gln Leu Arg Ile  
145 150 155 160

Asp Glu Ile Asn Gly Leu Leu Lys Ser Gly Arg Glu Asn Asn Lys Pro  
165 170 175

Leu Thr Asp Glu Arg Arg Ala Leu Leu Glu Ser Thr Ser Arg Ala Ala  
 180 185 190

Ala Gly Pro Ser Ile Phe His Pro Gly Gly Val Pro Gly Lys Cys Thr  
 195 200 205

Gln Phe Ala Leu Xaa  
 210

<210> 168

<211> 75

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (75)

<223> Xaa equals stop translation

<400> 168

Met Phe Thr Ser Phe Gly Leu Ala Ser Pro Arg Ile Leu Phe Cys Phe  
 1 5 10 15

Cys Phe Phe Asp Leu Gly Phe Ile Phe Phe Cys Val Leu Tyr Tyr Ile  
 20 25 30

Val Lys Gly Ile Leu Ala Glu Thr Leu Val Phe Gly Ala Arg Gly Glu  
 35 40 45

Gln Glu Cys Trp Ala Val Tyr Phe Arg Trp Arg Thr His Leu Gln Thr  
 50 55 60

Phe Gly Leu Phe Ser Phe Asn Cys Ser Val Xaa  
 65 70 75

<210> 169

<211> 48

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (48)

<223> Xaa equals stop translation

<400> 169

Met Phe Leu Cys Leu Phe Phe Phe Phe Phe Asn Ala Thr Gln Gly Asn  
 1 5 10 15

Ile Phe Ile Ser Phe Leu Ser Gly Leu Pro Gln Cys Ile Phe Ile Ser  
 20 25 30

Phe Glu Thr Lys Arg Phe Trp Lys Leu Phe Phe Cys Ser Phe Lys Xaa  
 35 40 45

<210> 170  
<211> 88  
<212> PRT  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (88)  
<223> Xaa equals stop translation

<400> 170  
Met Gly Leu His Leu Arg Pro Tyr Arg Val Gly Leu Leu Pro Asp Gly  
1 5 10 15  
Leu Leu Phe Leu Leu Leu Leu Met Leu Leu Ala Asp Pro Ala Leu  
20 25 30  
Pro Ala Gly Arg His Pro Pro Val Val Leu Val Pro Gly Asp Leu Gly  
35 40 45  
Asn Gln Leu Glu Ala Lys Leu Asp Lys Pro Thr Val Val His Tyr Leu  
50 55 60  
Cys Ser Lys Lys Thr Glu Ser Tyr Phe Thr Ile Trp Leu Asn Leu Glu  
65 70 75 80  
Leu Leu Leu Pro Val His His Xaa  
85

<210> 171  
<211> 42  
<212> PRT  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (42)  
<223> Xaa equals stop translation

<400> 171  
Met Ala Cys Glu Thr His Gly Val Leu Val Pro Ala His Leu Ser Gly  
1 5 10 15  
Leu Ile Thr Cys Leu Leu Ala Phe Trp Val Pro Ala Ser Cys Ile Gln  
20 25 30  
Arg Cys Ser Gly Ser Pro Leu Pro Leu Xaa  
35 40

<210> 172  
<211> 48  
<212> PRT  
<213> Homo sapiens



<220>  
 <221> SITE  
 <222> (48)  
 <223> Xaa equals stop translation

<400> 172  
 Met Gln Cys Phe Leu Phe Ser Ile Phe Leu Ile Thr Gly Leu Ala Glu  
     1                    5                    10                    15  
 Glu Phe Cys Glu Gln Leu Ser Ile Ser Leu Ala Glu Glu Glu Ile Gln  
                     20                    25                    30  
 Leu Ser Ser Thr Val Glu His Phe Cys Met Thr Ala Phe Ser Trp Xaa  
                     35                    40                    45

<210> 173  
 <211> 233  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (233)  
 <223> Xaa equals stop translation

<400> 173  
 Met Ala Ala Leu Ala Ala Ala Lys Lys Val Trp Ser Ala Arg Arg  
     1                    5                    10                    15  
 Leu Leu Val Leu Leu Phe Thr Pro Leu Ala Leu Leu Pro Val Val Phe  
                     20                    25                    30  
 Ala Leu Pro Pro Lys Glu Gly Arg Cys Leu Phe Val Ile Leu Leu Met  
                     35                    40                    45  
 Ala Val Tyr Trp Cys Thr Glu Ala Leu Pro Leu Ser Val Thr Ala Leu  
                     50                    55                    60  
 Leu Pro Ile Val Leu Phe Pro Phe Met Gly Ile Leu Pro Ser Asn Lys  
     65                    70                    75                    80  
 Val Cys Pro Gln Tyr Phe Leu Asp Thr Asn Phe Leu Phe Leu Ser Gly  
                     85                    90                    95  
 Leu Ile Met Ala Ser Ala Ile Glu Glu Trp Asn Leu His Arg Arg Ile  
                     100                    105                    110  
 Ala Leu Lys Ile Leu Met Leu Val Gly Val Gln Pro Ala Arg Leu Ile  
                     115                    120                    125  
 Leu Gly Met Met Val Thr Thr Ser Phe Leu Ser Met Trp Leu Ser Asn  
     130                    135                    140  
 Thr Ala Ser Thr Ala Met Met Leu Pro Ile Ala Asn Ala Ile Leu Lys  
     145                    150                    155                    160

Ser Leu Phe Gly Gln Lys Glu Val Arg Lys Asp Pro Ser Gln Glu Ser  
                                   165                                  170                                  175

Glu Glu Asn Thr Gly Ile Glu Pro Asn Thr Phe Leu Ser Glu Glu Arg  
                                   180                                  185                                  190

Leu Lys Leu Gln Ala Pro Leu Val Ile Arg Leu Gly Gln Ile Thr Glu  
                                   195                                  200                                  205

Ser Gly Gln Trp Asn Met Ser Gly Asn Asp Val Cys Asn Phe Arg Val  
                                   210                                  215                                  220

Leu Ser Phe Leu Pro Gly Gly Met Xaa .  
 225                                  230

<210> 174

<211> 45

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (45)

<223> Xaa equals stop translation

<400> 174

Met Gly Thr Ile Phe Gly Tyr Leu His Cys Val Lys Cys Tyr Val Leu  
   1                                  5                                  10                                  15

Tyr Phe Ile Phe Ile Leu Ile Thr Ala Val Tyr His Ser Phe Tyr Tyr  
                                   20                                  25                                  30

Pro His Tyr Arg Gly Lys Ala Leu Ile Ser Gly Thr Xaa  
                                   35                                  40                                  45

<210> 175

<211> 85

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (77)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (85)

<223> Xaa equals stop translation

<400> 175

Met Val Trp Phe Leu Phe Leu Val Phe Ile Phe Leu Lys Val Lys Gly  
   1                                  5                                  10                                  15

Asp Phe Phe Pro Pro Phe Leu Ile Cys Asn Leu Phe Cys Ile Trp Met  
                                   20                                  25                                  30

Ile Thr Gly Val Ser His Arg Leu Gln Pro Gln Ile Leu Phe Ser Arg  
 35 40 45

His Lys His Asn Gln Glu Ile Ile Leu Gln Met Val Ser Phe Ser Cys  
 50 55 60

Cys Val Phe Phe Pro Met Ile Arg Glu Val Lys Ser Xaa Leu Gly Cys  
 65 70 75 80

Ile Lys Met Ser Xaa  
 85

<210> 176

<211> 66

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (66)

<223> Xaa equals stop translation

<400> 176

Met Trp Val Leu Leu Ser Cys Pro Leu Pro Pro Leu Cys Leu Pro Ala  
 1 5 10 15

Ser Ala Val Pro Gly Gln Cys Leu Gly Gly Gln Trp Ser Gly His Gln  
 20 25 30

Leu Arg Leu Arg Gly Arg Gly Trp His Cys Arg Cys His Cys Arg Ala  
 35 40 45

Trp Ala Ala Asp Met Gly Arg Gly Leu His Ser Cys Gln Leu Leu Ser  
 50 55 60

Arg Xaa  
 65

<210> 177

<211> 55

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (55)

<223> Xaa equals stop translation

<400> 177

Met Leu Leu Leu Cys Ile Leu Leu Ile Phe Cys Val Val Gly Leu Ser  
 1 5 10 15

Val Val Gly Arg Arg Val Leu Lys Ser Thr Thr Ile Ile Val Tyr Leu  
 20 25 30

Ser Ile Thr Pro Phe Ser Ser Phe Ser Ser Ile Ser His Ile Phe Gln

35

40

45

Leu Leu Ile Gly Ala His Xaa  
 50 55

&lt;210&gt; 178

&lt;211&gt; 83

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (4)

&lt;223&gt; Xaa equals any of the naturally occurring L-amino acids

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (83)

&lt;223&gt; Xaa equals stop translation

&lt;400&gt; 178

Met Cys Val Xaa Leu Ser Phe Cys Pro Phe Leu Ser Ser Ala Leu Pro  
 1 5 10 15

Ala Ser His Thr Gln Phe Tyr Met Pro Arg Gly Ala Lys Phe Gly Thr  
 20 25 30

Phe Thr Leu Gln Ala Ser Val Ser Pro Leu Glu Glu Lys Thr His Ser  
 35 40 45

Phe Thr His Pro Gly Ile Gly Gly Lys Leu Leu Gly His Gln Asp Pro  
 50 55 60

Gly Ala Pro Gly Pro Ser Trp Asn Ile Arg Ser Thr Trp Ser Thr Arg  
 65 70 75 80

Ser Leu Xaa

&lt;210&gt; 179

&lt;211&gt; 330

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (38)

&lt;223&gt; Xaa equals any of the naturally occurring L-amino acids

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (247)

&lt;223&gt; Xaa equals any of the naturally occurring L-amino acids

&lt;400&gt; 179

Met Ser Pro Leu Ser Ala Ala Arg Ala Ala Leu Arg Val Tyr Ala Val  
 1 5 10 15

Gly Ala Ala Val Ile Leu Ala Gln Leu Leu Arg Arg Cys Arg Gly Gly  
                   20                                  25                                  30  
 Phe Leu Glu Pro Val Xaa Pro Pro Arg Pro Asp Arg Val Ala Ile Val  
                   35                                  40                                  45  
 Thr Gly Gly Thr Asp Gly Ile Gly Tyr Ser Thr Ala Lys His Leu Ala  
                   50                                  55                                  60  
 Arg Leu Gly Met His Val Ile Ile Ala Gly Asn Asn Asp Ser Lys Ala  
                   65                                  70                                  75                                  80  
 Lys Gln Val Val Ser Lys Ile Lys Glu Glu Thr Leu Asn Asp Lys Val  
                                   85                                  90                                  95  
 Glu Phe Leu Tyr Cys Asp Leu Ala Ser Met Thr Ser Ile Arg Gln Phe  
                                   100                                  105                                  110  
 Val Gln Lys Phe Lys Met Lys Lys Ile Pro Leu His Val Leu Ile Asn  
                                   115                                  120                                  125  
 Asn Ala Gly Val Met Met Val Pro Gln Arg Lys Thr Arg Asp Gly Phe  
                                   130                                  135                                  140  
 Glu Glu His Phe Gly Leu Asn Tyr Leu Gly His Phe Leu Leu Thr Asn  
                                   145                                  150                                  155                                  160  
 Leu Leu Leu Asp Thr Leu Lys Glu Ser Gly Ser Pro Gly His Ser Ala  
                                   165                                  170                                  175  
 Arg Val Val Thr Val Ser Ser Ala Thr His Tyr Val Ala Glu Leu Asn  
                                   180                                  185                                  190  
 Met Asp Asp Leu Gln Ser Ser Ala Cys Tyr Ser Pro His Ala Ala Tyr  
                                   195                                  200                                  205

---

Ala Gln Ser Lys Leu Ala Leu Val Leu Phe Thr Tyr His Leu Gln Arg  
                   210                                  215                                  220  
 Leu Leu Ala Ala Glu Gly Ser His Val Thr Ala Asn Val Val Asp Pro  
                   225                                  230                                  235                                  240  
 Gly Val Val Asn Thr Asp Xaa Tyr Lys His Val Phe Trp Ala Thr Arg  
                                   245                                  250                                  255  
 Leu Ala Lys Lys Leu Leu Gly Trp Leu Leu Phe Lys Thr Pro Asp Glu  
                                   260                                  265                                  270  
 Gly Ala Trp Thr Ser Ile Tyr Ala Ala Val Thr Pro Glu Leu Glu Gly  
                                   275                                  280                                  285  
 Val Gly Gly Arg Tyr Leu Tyr Asn Glu Lys Glu Thr Lys Ser Leu His  
                                   290                                  295                                  300  
 Val Thr Tyr Asn Gln Lys Leu Gln Gln Gln Leu Trp Ser Lys Ser Cys  
                                   305                                  310                                  315                                  320  
 Glu Met Thr Gly Val Leu Asp Val Thr Leu

325

330

<210> 180  
<211> 41  
<212> PRT  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (41)  
<223> Xaa equals stop translation

<400> 180  
Met Ile Ala Cys Gln Tyr Ile Ser Leu Ala Ile Met Leu Ala Phe Val  
1 5 10 15

Arg Trp Ala Ala Phe Leu Leu Phe Pro Phe Leu Cys Gly Asp Asn Gly  
20 25 30

Gly Asn Ile Gln Gln Lys Tyr Val Xaa  
35 40

<210> 181  
<211> 52  
<212> PRT  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (52)  
<223> Xaa equals stop translation

<400> 181  
Met Ala Asn Ala Met Ala Tyr Leu Ser Ile Phe Leu Cys Gly Ala Ser  
1 5 10 15

Ser Ser Pro Cys Asp Cys Ala Leu Leu Val Pro Val Ser Leu Phe Arg  
20 25 30

Gly Arg Lys Val Ala Asn Phe Lys Asn Gln Asn Ser Asp Val Thr Ser  
35 40 45

Gly Asn Ala Xaa  
50

<210> 182  
<211> 55  
<212> PRT  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (55)  
<223> Xaa equals stop translation

<400> 182

Met Gln Gln Ile Cys Ser Cys Leu Gly Ala Phe Ala Leu Leu Phe Phe  
 1 5 10 15  
 Trp Pro Gly His Phe Thr Ser Thr Phe Ser Ile Phe Tyr Asp Phe Leu  
 20 25 30  
 Pro Ile Phe Gly Ser Leu Phe Lys Cys His Pro Ser Lys Arg Pro Ser  
 35 40 45  
 Lys Leu Pro Tyr Leu Lys Xaa  
 50 55

<210> 183  
 <211> 62  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (62)  
 <223> Xaa equals stop translation

<400> 183  
 Met Arg Leu Leu Leu Glu Trp Arg Val Tyr Leu Arg Leu Thr Cys Ala  
 1 5 10 15  
 Thr Lys Asp Gly Met Ala Arg Glu Cys Pro Thr Thr Trp Leu Ser Pro  
 20 25 30  
 Pro Ala Lys Pro Asp Phe Ala Gln Arg His Ser Val Lys Pro Thr Ala  
 35 40 45  
 Leu Gln Gly Gly Arg Trp Ser Arg Leu Gly Ala Ser Pro Xaa  
 50 55 60

<210> 184  
 <211> 148  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (148)  
 <223> Xaa equals stop translation

<400> 184  
 Met Leu Gly Leu Pro Trp Lys Gly Gly Leu Ser Trp Ala Leu Leu Leu  
 1 5 10 15  
 Leu Leu Leu Gly Ser Gln Ile Leu Leu Ile Tyr Ala Trp His Phe His  
 20 25 30  
 Glu Gln Arg Asp Cys Asp Glu His Asn Val Met Ala Arg Tyr Leu Pro  
 35 40 45  
 Ala Thr Val Glu Phe Ala Val His Thr Phe Asn Gln Gln Ser Lys Asp  
 50 55 60

Tyr Tyr Ala Tyr Arg Leu Gly His Ile Leu Asn Ser Trp Lys Glu Gln  
65 70 75 80

Val Glu Ser Lys Thr Val Phe Ser Met Glu Leu Leu Leu Gly Arg Thr  
85 90 95

Arg Cys Gly Lys Phe Glu Asp Asp Ile Asp Asn Cys His Phe Gln Glu  
100 105 110

Ser Thr Glu Leu Asn Asn Thr Phe Thr Cys Phe Phe Thr Ile Ser Thr  
115 120 125

Arg Pro Trp Met Thr Gln Phe Ser Leu Leu Asn Lys Thr Cys Leu Glu  
130 135 140

Gly Phe His Xaa  
145

<210> 185

<211> 161

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (146)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (151)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (161)

<223> Xaa equals stop translation

<400> 185

Met Arg Leu Leu Cys Gly Leu Trp Leu Trp Leu Ser Leu Leu Lys Val  
1 5 10 15

Leu Gln Ala Gln Thr Pro Thr Pro Leu Pro Leu Pro Pro Pro Met Gln  
20 25 30

Ser Phe Gln Gly Asn Gln Phe Gln Gly Glu Trp Phe Val Leu Gly Leu  
35 40 45

Ala Gly Asn Ser Phe Arg Pro Glu His Arg Ala Leu Leu Asn Ala Phe  
50 55 60

Thr Ala Thr Phe Glu Leu Ser Asp Asp Gly Arg Phe Glu Val Trp Asn  
65 70 75 80

Ala Met Thr Arg Gly Gln His Cys Asp Thr Trp Ser Tyr Val Leu Ile  
85 90 95



131

Pro Ala Ala Gln Pro Gly Gln Phe Thr Val Asp His Gly Val Gly Arg  
                   100                                  105                                  110

Ser Trp Leu Leu Pro Pro Gly Thr Leu Asp Gln Phe Ile Cys Leu Gly  
                   115                                  120                                  125

Arg Ala Gln Gly Leu Ser Asp Asp Asn Ile Val Phe Pro Asp Val Thr  
                   130                                  135                                  140

Gly Xaa Ala Leu Asp Leu Xaa Ser Leu Pro Trp Val Ala Ala Pro Ala  
                   145                                  150                                  155                                  160

Xaa

<210> 186  
 <211> 122  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (122)  
 <223> Xaa equals stop translation

<400> 186  
 Met Met Leu Pro Gln Trp Leu Leu Leu Leu Phe Leu Leu Phe Phe Phe  
       1                                  5                                  10                                  15

Leu Phe Leu Leu Thr Arg Gly Ser Leu Ser Pro Thr Lys Tyr Asn Leu  
                   20                                  25                                  30

Leu Glu Leu Lys Glu Ser Cys Ile Arg Asn Gln Asp Cys Glu Thr Gly  
                   35                                  40                                  45

Cys Cys Gln Arg Ala Pro Asp Asn Cys Glu Ser His Cys Ala Glu Lys  
       50                                  55                                  60

Gly Ser Glu Gly Ser Leu Cys Gln Thr Gln Val Phe Phe Gly Gln Tyr  
       65                                  70                                  75                                  80

Arg Ala Cys Pro Cys Leu Arg Asn Leu Thr Cys Ile Tyr Ser Lys Asn  
                   85                                  90                                  95

Glu Lys Trp Leu Ser Ile Ala Tyr Gly Arg Cys Gln Lys Ile Gly Arg  
                   100                                  105                                  110

Gln Lys Leu Ala Lys Lys Met Phe Phe Xaa  
                   115                                  120

<210> 187  
 <211> 163  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE

&lt;222&gt; (163)

&lt;223&gt; Xaa equals stop translation

&lt;400&gt; 187

Met Thr Ser Asn Phe Pro Phe Cys Thr Leu Ile Leu Gly Ile Ala Gln  
 1 5 10 15

Ala Gln Ala Cys Pro Gly Cys Pro Gly Asp Trp Pro Gly Leu Gly Ser  
 20 25 30

Gly Val Gly Glu Gly Leu His His Ile Arg Thr Cys Arg Thr Pro Ile  
 35 40 45

Pro Cys Ser Pro Pro Ala Pro Ala Ala Ala Cys Leu Gly Ser Gly His  
 50 55 60

Ala Arg Leu Pro Cys Val Leu Arg Leu Trp Pro Val Pro Ala Asn Leu  
 65 70 75 80

Ser Ser Pro Phe Arg Leu Glu Ala Leu His Cys Ser Phe Trp Ser Ser  
 85 90 95

Pro Leu Leu Pro Ala Pro His Leu Ala Phe Phe Gly Phe Arg Asp Leu  
 100 105 110

Leu Thr Asp Phe Leu Leu Ala Ala Cys Leu Leu Thr Phe Gln Lys Thr  
 115 120 125

Pro Leu Glu Leu Pro Met Ala Val Val His Leu Leu Val Ala Thr Pro  
 130 135 140

Cys Tyr Gln Met Leu Asp Asn Leu Pro Leu Pro Ser Ala Ala Ala Asn  
 145 150 155 160

Trp Cys Xaa

&lt;210&gt; 188

&lt;211&gt; 51

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (51)

&lt;223&gt; Xaa equals stop translation

&lt;400&gt; 188

Met Pro Gly Ile Leu Ala Gly Ile Pro Val Lys Asp Leu Cys Leu Ser  
 1 5 10 15

Leu Leu Gln Gly Phe Arg Leu Leu Leu Cys Val Cys Pro Gly Trp  
 20 25 30

Leu Ser Gly Trp Met Gly Gly Gln Lys Gly Ser Pro Arg Ile Val Asp  
 35 40 45

Ile Gly Xaa

50

<210> 189  
<211> 65  
<212> PRT  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (65)  
<223> Xaa equals stop translation

<400> 189  
Met Tyr Leu Tyr Leu Gly Val Phe Phe His Leu Ile Tyr Pro Gly Ala  
1 5 10 15  
Leu Ser Ile Thr Thr Leu Gly Lys His Ser His Pro Phe Phe Thr Ala  
20 25 30  
Glu Gln Asn Ser Thr Val Trp Met Glu His Thr Leu Phe His Gln Ser  
35 40 45  
Pro Val Ala Ser His Leu Val Cys Phe Gln Ser Phe Ala Phe Ser Glu  
50 55 60

Xaa  
65

<210> 190  
<211> 47  
<212> PRT  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (47)  
<223> Xaa equals stop translation

<400> 190  
Met Thr Leu Ser Leu Gln Leu Ala Glu Leu Val His Phe Val Cys Ala  
1 5 10 15  
Phe Gln Ser Gln Trp Thr Gly Val Tyr Pro Met Met Pro Pro Leu Lys  
20 25 30  
Pro Thr Glu Pro Leu Cys Phe Ala Cys Val Pro Cys Arg Val Xaa  
35 40 45

<210> 191  
<211> 144  
<212> PRT  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (144)

<223> Xaa equals stop translation

<400> 191

Met Ser Pro Phe His Leu Leu Gly Leu Lys Val Phe Leu Thr Trp Ala  
1 5 10 15

Leu Thr Leu Ala Gln Ile Cys Leu Tyr Phe Phe Glu Val Gln Pro Leu  
20 25 30

Gly Leu Leu Ala Leu Asn Phe Phe Cys Thr Ala Thr Ala Gly Leu Lys  
35 40 45

Glu Leu Cys Met His Pro Pro Ser Leu Ala Phe Thr Pro Glu Phe His  
50 55 60

Thr Ser Leu Ser Pro Leu Ala Ile Pro Ser Phe Cys Gly Thr Ser Val  
65 70 75 80

Ser Leu Ser Asn Ser His Thr Ile Pro Leu Ser Leu Tyr Leu Pro Phe  
85 90 95

Pro Ser Lys Ser Arg Met Pro Asp Thr Leu His Leu Leu Val His Ser  
100 105 110

Leu Pro Leu Val His Ser Gln Val Leu Pro Val Lys Asp Val Thr Ile  
115 120 125

Glu Trp Pro Leu Cys Gln Arg Cys Leu Gly Ser Thr Cys His Gln Xaa  
130 135 140

<210> 192

<211> 81

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (76)

<223> Xaa equals any of the naturally occurring L-amino acids

<220>

<221> SITE

<222> (81)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 192

Met Phe Cys Phe Ser Ser Ile Phe Cys Ser His Glu His Thr His Leu  
1 5 10 15

Pro Gly Thr Phe Trp Leu Phe Leu Phe Leu Phe Leu Ile Leu Pro Pro  
20 25 30

Ser Cys Pro Cys Phe Leu Pro Phe Ser Leu Ala Ile Glu Thr Val Arg  
35 40 45

135

Trp Pro Cys Trp His His Pro Thr Ser Phe Glu Leu Cys Tyr Pro Gly  
50 55 60

Thr Ser Ile Tyr Tyr Ala Ser Arg Gly Gly Pro Xaa Pro Asn Ser Glu  
65 70 75 80

Xaa

<210> 193

<211> 45

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (45)

<223> Xaa equals stop translation

<400> 193

Met Thr Tyr Leu Phe Cys Ser Ser Ile Ser Leu Leu Leu Lys Val  
1 5 10 15

His Ser Ser Gly His Gln Asp Ile Arg Lys Ala Lys Ser Lys Val Pro  
20 25 30

Arg Leu Leu Ile Ile Gln Cys Pro Gln Gln Arg Glu Xaa  
35 40 45

<210> 194

<211> 42

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (42)

<223> Xaa equals stop translation

<400> 194

Met Pro Thr Ile Trp Val Lys Leu Cys Leu Leu Gln Val Cys His Gly  
1 5 10 15

Leu Phe Pro Leu Leu Lys His Trp Ser Gln Pro Met Pro Leu Cys Val  
20 25 30

Thr Leu Ala Pro Val Ser Tyr Trp Leu Xaa  
35 40

<210> 195

<211> 260

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

&lt;222&gt; (260)

&lt;223&gt; Xaa equals stop translation

&lt;400&gt; 195

```

Met Gly Thr Ala Ala Leu Gly Pro Val Trp Ala Ala Leu Leu Leu Phe
 1           5           10           15
Leu Leu Met Cys Glu Ile Pro Met Val Glu Leu Thr Phe Asp Arg Ala
          20           25           30
Val Ala Ser Asp Cys Gln Arg Cys Asp Ser Glu Asp Pro Leu Asp
          35           40           45
Pro Ala His Val Ser Ser Ala Ser Ser Ser Gly Arg Pro His Ala Leu
          50           55           60
Pro Glu Ile Arg Pro Tyr Ile Asn Ile Thr Ile Leu Lys Gly Asp Lys
          65           70           75           80
Gly Asp Pro Gly Pro Met Gly Leu Pro Gly Tyr Met Gly Arg Glu Gly
          85           90           95
Pro Gln Gly Glu Pro Gly Pro Gln Gly Ser Lys Gly Asp Lys Gly Glu
          100          105          110
Met Gly Ser Pro Gly Ala Pro Cys Gln Lys Arg Phe Phe Ala Phe Ser
          115          120          125
Val Gly Arg Lys Thr Ala Leu His Ser Gly Glu Asp Phe Gln Thr Leu
          130          135          140
Leu Phe Glu Arg Val Phe Val Asn Leu Asp Gly Cys Phe Asp Met Ala
          145          150          155          160
Thr Gly Gln Phe Ala Ala Pro Leu Arg Gly Ile Tyr Phe Phe Ser Leu
          165          170          175
Asn Val His Ser Trp Asn Tyr Lys Glu Thr Tyr Val His Ile Met His
          180          185          190
Asn Gln Lys Glu Ala Val Ile Leu Tyr Ala Gln Pro Ser Glu Arg Ser
          195          200          205
Ile Met Gln Ser Gln Ser Val Met Leu Asp Leu Ala Tyr Gly Asp Arg
          210          215          220
Val Trp Val Arg Leu Phe Lys Arg Gln Arg Glu Asn Ala Ile Tyr Ser
          225          230          235          240
Asn Asp Phe Asp Thr Tyr Ile Thr Phe Ser Gly His Leu Ile Lys Ala
          245          250          255
Glu Asp Asp Xaa
          260

```

&lt;210&gt; 196

&lt;211&gt; 117

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 196

Met Leu Gly His Cys Cys Tyr Phe Trp Gln Val Trp Pro Ala Ser Glu  
 1 5 10 15

Ala Leu Ala Ala Gly Pro Thr Pro Ser Thr Gly Ser Ser Ser Pro Ser  
 20 25 30

Trp Lys Gln His Ile Gly Thr Ser Leu Gln Lys Thr Arg Gly Ser Leu  
 35 40 45

Pro Thr Thr Thr Leu Thr Ser Gly Ala Gly Gln Ser Thr Ser Thr Gly  
 50 55 60

Lys Asn Pro Ala Ala Gly Arg Ser Leu Glu Gly Ala Leu Pro Ala Gly  
 65 70 75 80

Val Trp Pro Cys Phe Ala Gln Ser Pro Cys Thr Gly Gly Gln Gln Thr  
 85 90 95

Pro Ser Ser Thr Gly Leu Arg Ser Cys Leu Val Arg Ser Pro Ala Thr  
 100 105 110

Trp Trp Arg Thr Pro  
 115

&lt;210&gt; 197

&lt;211&gt; 698

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 197

Met Leu Pro Ala Arg Leu Pro Phe Arg Leu Leu Ser Leu Phe Leu Arg  
 1 5 10 15

Gly Ser Ala Pro Thr Ala Ala Arg His Gly Leu Arg Glu Pro Leu Leu  
 20 25 30

Glu Arg Arg Cys Ala Ala Ala Ser Ser Phe Gln His Ser Ser Ser Leu  
 35 40 45

Gly Arg Glu Leu Pro Tyr Asp Pro Val Asp Thr Glu Gly Phe Gly Glu  
 50 55 60

Gly Gly Asp Met Gln Glu Arg Phe Leu Phe Pro Glu Tyr Ile Leu Asp  
 65 70 75 80

Pro Glu Pro Gln Pro Thr Arg Glu Lys Gln Leu Gln Glu Leu Gln Gln  
 85 90 95

Gln Gln Glu Glu Glu Arg Gln Arg Gln Gln Arg Arg Glu Glu Arg  
 100 105 110

Arg Gln Gln Asn Leu Arg Ala Arg Ser Arg Glu His Pro Val Val Gly  
 115 120 125

His Pro Asp Pro Ala Leu Pro Pro Ser Gly Val Asn Cys Ser Gly Cys

130	135	140
Gly Ala Glu Leu His Cys Gln Asp Ala Gly Val Pro Gly Tyr Leu Pro		
145	150	155 160
Arg Glu Lys Phe Leu Arg Thr Ala Glu Ala Asp Gly Gly Leu Ala Arg		
	165	170 175
Thr Val Cys Gln Arg Cys Trp Leu Leu Ser His His Arg Arg Ala Leu		
	180	185 190
Arg Leu Gln Val Ser Arg Glu Gln Tyr Leu Glu Leu Val Ser Ala Ala		
	195	200 205
Leu Arg Arg Pro Gly Pro Ser Leu Val Leu Tyr Met Val Asp Leu Leu		
	210	215 220
Asp Leu Pro Asp Ala Leu Leu Pro Asp Leu Pro Ala Leu Val Gly Pro		
	225	230 235 240
Lys Gln Leu Ile Val Leu Gly Asn Lys Val Asp Leu Leu Pro Gln Asp		
	245	250 255
Ala Pro Gly Tyr Arg Gln Arg Leu Arg Glu Arg Leu Trp Glu Asp Cys		
	260	265 270
Ala Arg Ala Gly Leu Leu Leu Ala Pro Gly His Gln Gly Pro Gln Arg		
	275	280 285
Pro Val Lys Asp Glu Pro Gln Asp Gly Glu Asn Pro Asn Pro Pro Asn		
	290	295 300
Trp Ser Arg Thr Val Val Arg Asp Val Arg Leu Ile Ser Ala Lys Thr		
	305	310 315 320
Gly Tyr Gly Val Glu Glu Leu Ile Ser Ala Leu Gln Arg Ser Trp Arg		
	325	330 335
Tyr Arg Gly Asp Val Tyr Leu Val Gly Ala Thr Asn Ala Gly Lys Ser		
	340	345 350
Thr Leu Phe Asn Thr Leu Leu Glu Ser Asp Tyr Cys Thr Ala Lys Gly		
	355	360 365
Ser Asp Ala Ile Asp Arg Ala Thr Ile Ser Pro Trp Pro Gly Thr Thr		
	370	375 380
Leu Asn Leu Leu Lys Phe Pro Ile Cys Asn Pro Thr Pro Tyr Arg Met		
	385	390 395 400
Phe Lys Arg His Gln Arg Leu Lys Lys Asp Ser Thr Gln Ala Glu Glu		
	405	410 415
Asp Leu Ser Glu Gln Glu Gln Asn Gln Leu Asn Val Leu Lys Lys His		
	420	425 430
Gly Tyr Val Val Gly Arg Val Gly Arg Thr Phe Leu Tyr Ser Glu Glu		
	435	440 445



Gln Lys Asp Asn Ile Pro Phe Glu Phe Asp Ala Asp Ser Leu Ala Phe  
 450 455 460  
 Asp Met Glu Asn Asp Pro Val Met Gly Thr His Lys Ser Thr Lys Gln  
 465 470 475 480  
 Val Glu Leu Thr Ala Gln Asp Val Lys Asp Ala His Trp Phe Tyr Asp  
 485 490 495  
 Thr Pro Gly Ile Thr Lys Glu Asn Cys Ile Leu Asn Leu Leu Thr Glu  
 500 505 510  
 Lys Glu Val Asn Ile Val Leu Pro Thr Gln Ser Ile Val Pro Arg Thr  
 515 520 525  
 Phe Val Leu Lys Pro Gly Met Val Leu Phe Leu Gly Ala Ile Gly Arg  
 530 535 540  
 Ile Asp Phe Leu Gln Gly Asn Gln Ser Ala Trp Phe Thr Val Val Ala  
 545 550 555 560  
 Ser Asn Ile Leu Pro Val His Ile Thr Ser Leu Asp Arg Ala Asp Ala  
 565 570 575  
 Leu Tyr Gln Lys His Ala Gly His Thr Leu Leu Gln Ile Pro Met Gly  
 580 585 590  
 Gly Lys Glu Arg Met Ala Gly Phe Pro Pro Leu Val Ala Glu Asp Ile  
 595 600 605  
 Met Leu Lys Glu Gly Leu Gly Ala Ser Glu Ala Val Ala Asp Ile Lys  
 610 615 620  
 Phe Ser Ser Ala Gly Trp Val Ser Val Thr Pro Asn Phe Lys Asp Arg  
 625 630 635 640

---

Leu His Leu Arg Gly Tyr Thr Pro Glu Gly Thr Val Leu Thr Val Arg  
 645 650 655  
 Pro Pro Leu Leu Pro Tyr Ile Val Asn Ile Lys Gly Gln Arg Ile Lys  
 660 665 670  
 Lys Ser Val Ala Tyr Lys Thr Lys Lys Pro Pro Ser Leu Met Tyr Asn  
 675 680 685  
 Val Arg Lys Lys Lys Gly Lys Ile Asn Val  
 690 695

<210> 198  
 <211> 348  
 <212> PRT  
 <213> Homo sapiens

<400> 198  
 Met Asn Met Thr Gln Ala Arg Val Leu Val Ala Ala Val Val Gly Leu  
 1 5 10 15  
 Val Ala Val Leu Leu Tyr Ala Ser Ile His Lys Ile Glu Glu Gly His

140

20	25	30
Leu Ala Val Tyr Tyr Arg Gly Gly Ala Leu Leu Thr Ser Pro Ser Gly 35 40 45		
Pro Gly Tyr His Ile Met Leu Pro Phe Ile Thr Thr Phe Arg Ser Val 50 55 60		
Gln Thr Thr Leu Gln Thr Asp Glu Val Lys Asn Val Pro Cys Gly Thr 65 70 75 80		
Ser Gly Gly Val Met Ile Tyr Ile Asp Arg Ile Glu Val Val Asn Met 85 90 95		
Leu Ala Pro Tyr Ala Val Phe Asp Ile Val Arg Asn Tyr Thr Ala Asp 100 105 110		
Tyr Asp Lys Thr Leu Ile Phe Asn Lys Ile His His Glu Leu Asn Gln 115 120 125		
Phe Cys Ser Ala His Thr Leu Gln Glu Val Tyr Ile Glu Leu Phe Asp 130 135 140		
Gln Ile Asp Glu Asn Leu Lys Gln Ala Leu Gln Lys Asp Leu Asn Leu 145 150 155 160		
Met Ala Pro Gly Leu Thr Ile Gln Ala Val Arg Val Thr Lys Pro Lys 165 170 175		
Ile Pro Glu Ala Ile Arg Arg Asn Phe Glu Leu Met Glu Ala Glu Lys 180 185 190		
Thr Lys Leu Leu Ile Ala Ala Gln Lys Gln Lys Val Val Glu Lys Glu 195 200 205		
Ala Glu Thr Glu Arg Lys Lys Ala Val Ile Glu Ala Glu Lys Ile Ala 210 215 220		
Gln Val Ala Lys Ile Arg Phe Gln Gln Lys Val Met Glu Lys Glu Thr 225 230 235 240		
Glu Lys Arg Ile Ser Glu Ile Glu Asp Ala Ala Phe Leu Ala Arg Glu 245 250 255		
Lys Ala Lys Ala Asp Ala Glu Tyr Tyr Ala Ala His Lys Tyr Ala Thr 260 265 270		
Ser Asn Lys His Lys Leu Thr Pro Glu Tyr Leu Glu Leu Lys Lys Tyr 275 280 285		
Gln Ala Ile Ala Ser Asn Ser Lys Ile Tyr Phe Gly Ser Asn Ile Pro 290 295 300		
Asn Met Phe Val Asp Ser Ser Cys Ala Leu Lys Tyr Ser Asp Ile Arg 305 310 315 320		
Thr Gly Arg Glu Ser Ser Leu Pro Ser Lys Glu Ala Leu Glu Pro Ser 325 330 335		

Gly Glu Asn Val Ile Gln Asn Lys Glu Ser Thr Gly  
 340 345

<210> 199

<211> 401

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (307)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 199

Met Met Gly Leu Gly Asn Gly Arg Arg Ser Met Lys Ser Pro Pro Leu  
 1 5 10 15

Val Leu Ala Ala Leu Val Ala Cys Ile Ile Val Leu Gly Phe Asn Tyr  
 20 25 30

Trp Ile Ala Ser Ser Arg Ser Val Asp Leu Gln Thr Arg Ile Met Glu  
 35 40 45

Leu Glu Gly Arg Val Arg Arg Arg Ala Ala Glu Arg Gly Ala Val Glu  
 50 55 60

Leu Lys Lys Asn Glu Phe Gln Gly Glu Leu Glu Lys Gln Arg Glu Gln  
 65 70 75 80

Leu Asp Lys Ile Gln Ser Ser His Asn Phe Gln Leu Glu Ser Val Asn  
 85 90 95

Lys Leu Tyr Gln Asp Glu Lys Ala Val Leu Val Asn Asn Ile Thr Thr  
 100 105 110

Gly Glu Arg Leu Ile Arg Val Leu Gln Asp Gln Leu Lys Thr Leu Gln  
 115 120 125

Arg Asn Tyr Gly Arg Leu Gln Gln Asp Val Leu Gln Phe Gln Lys Asn  
 130 135 140

Gln Thr Asn Leu Glu Arg Lys Phe Ser Tyr Asp Leu Ser Gln Cys Ile  
 145 150 155 160

Asn Gln Met Lys Glu Val Lys Glu Gln Cys Glu Glu Arg Ile Glu Glu  
 165 170 175

Val Thr Lys Lys Gly Asn Glu Ala Val Ala Ser Arg Asp Leu Ser Glu  
 180 185 190

Asn Asn Asp Gln Arg Gln Gln Leu Gln Ala Leu Ser Glu Pro Gln Pro  
 195 200 205

Arg Leu Gln Ala Ala Gly Leu Pro His Thr Glu Val Pro Gln Gly Lys  
 210 215 220

Gly Asn Val Leu Gly Asn Ser Lys Ser Gln Thr Pro Ala Pro Ser Ser  
 225 230 235 240

```
<210> 200
<211> 324
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SITE
<222> (3)
<223> Xaa equals any of the naturally occurring L-amino acids
```

```
<400> 200
Met Glu Xaa Ala Lys Val Tyr Val Ala Lys Val Asp Cys Thr Ala His
  1                      5              10              15
```

Ser Asp Val Cys Ser Ala Gln Gly Val Arg Gly Tyr Pro Thr Leu Lys  
20 25 30

Leu Phe Lys Pro Gly Gln Glu Ala Val Lys Tyr Gln Gly Pro Arg Asp  
35 40 45

Phe Gln Thr Leu Glu Asn Trp Met Leu Gln Thr Leu Asn Glu Glu Pro  
50 55 60

Val Thr Pro Glu Pro Glu Val Glu Pro Pro Ser Ala Pro Glu Leu Lys

65	70	75	80
Gln Gly Leu Tyr Glu Leu Ser Ala Ser Asn Phe Glu Leu His Val Ala	85	90	95
Gln Gly Asp His Phe Ile Lys Phe Phe Ala Pro Trp Cys Gly His Cys	100	105	110
Lys Ala Leu Ala Pro Thr Trp Glu Gln Leu Ala Leu Gly Leu Glu His	115	120	125
Ser Glu Thr Val Lys Ile Gly Lys Val Asp Cys Thr Gln His Tyr Glu	130	135	140
Leu Cys Ser Gly Asn Gln Val Arg Gly Tyr Pro Thr Leu Leu Trp Phe	145	150	155
Arg Asp Gly Lys Lys Val Asp Gln Tyr Lys Gly Lys Arg Asp Leu Glu	165	170	175
Ser Leu Arg Glu Tyr Val Glu Ser Gln Leu Gln Arg Thr Glu Thr Gly	180	185	190
Ala Thr Glu Thr Val Thr Pro Ser Glu Ala Pro Val Leu Ala Ala Glu	195	200	205
Pro Glu Ala Asp Lys Gly Thr Val Leu Ala Leu Thr Glu Asn Asn Phe	210	215	220
Asp Asp Thr Ile Ala Glu Gly Ile Thr Phe Ile Lys Phe Tyr Ala Pro	225	230	235
Trp Cys Gly His Cys Lys Thr Leu Ala Pro Thr Trp Glu Glu Leu Ser	245	250	255
<del>Lys Lys Glu Phe Pro Gly Leu Ala Gly Val Lys Ile Ala Glu Val Asp</del>	<del>260</del>	<del>265</del>	<del>270</del>
Cys Thr Ala Glu Arg Asn Ile Cys Ser Lys Tyr Ser Val Arg Gly Tyr	275	280	285
Pro Thr Leu Leu Leu Phe Arg Gly Gly Lys Lys Val Ser Glu His Ser	290	295	300
Gly Gly Arg Asp Leu Asp Ser Leu His Arg Phe Val Leu Ser Gln Ala	305	310	315
			320
Lys Asp Glu Leu			

<210> 201  
 <211> 90  
 <212> PRT  
 <213> Homo sapiens

<400> 201  
 Met Ala Leu Phe Ser Cys Leu Leu Leu Leu Lys Gln Ser Asp Gly Ala  
 1 5 10 15

Ser Pro Val Leu Arg Ala Leu Ala Ala Ser Cys Leu Ala Ser Pro Ala  
20 25 30

Gly Cys Cys Gly Thr Arg Lys Ala Leu Asn Gly Asn Val Gly Glu Lys  
35 40 45

Val Gly Phe Thr Phe Met Ser Phe Gln Gly Cys Asp Pro Ser Ser Pro  
50 55 60

Gly Cys Leu Cys Cys Ser Leu Leu Pro Ser Asn Ser Gln Leu Val Phe  
65 70 75 80

Ile Ser Phe Leu Val Leu Ser Gly Leu Ala  
85 90

<210> 202

<211> 243

<212> PRT

<213> Homo sapiens

<400> 202

Met Arg Pro Gln Gly Pro Ala Ala Ser Pro Gln Arg Leu Arg Gly Leu  
1 5 10 15

Leu Leu Leu Leu Leu Leu Gln Leu Pro Ala Pro Ser Ser Ala Ser Glu  
20 25 30

Ile Pro Lys Gly Lys Gln Lys Ala Gln Leu Arg Gln Arg Glu Val Val  
35 40 45

Asp Leu Tyr Asn Gly Met Cys Leu Gln Gly Pro Ala Gly Val Pro Gly  
50 55 60

Arg Asp Gly Ser Pro Gly Ala Asn Gly Ile Pro Gly Thr Pro Gly Ile  
65 70 75 80

Pro Gly Arg Asp Gly Phe Lys Gly Glu Lys Gly Glu Cys Leu Arg Glu  
85 90 95

Ser Phe Glu Glu Ser Trp Thr Pro Asn Tyr Lys Gln Cys Ser Trp Ser  
100 105 110

Ser Leu Asn Tyr Gly Ile Asp Leu Gly Lys Ile Ala Glu Cys Thr Phe  
115 120 125

Thr Lys Met Arg Ser Asn Ser Ala Leu Arg Val Leu Phe Ser Gly Ser  
130 135 140

Leu Arg Leu Lys Cys Arg Asn Ala Cys Cys Glu Arg Trp Tyr Phe Thr  
145 150 155 160

Phe Asn Gly Ala Glu Cys Ser Gly Pro Leu Pro Ile Glu Ala Ile Ile  
165 170 175

Tyr Leu Asp Gln Gly Ser Pro Glu Met Asn Ser Thr Ile Asn Ile His  
180 185 190

Arg Thr Ser Ser Val Glu Gly Leu Cys Glu Gly Ile Gly Ala Gly Leu  
 195 200 205

Val Asp Val Ala Ile Trp Val Gly Thr Cys Ser Asp Tyr Pro Lys Gly  
 210 215 220

Asp Ala Ser Thr Gly Trp Asn Ser Val Ser Arg Ile Ile Ile Glu Glu  
 225 230 235 240

Leu Pro Lys

<210> 203

<211> 75

<212> PRT

<213> Homo sapiens

<400> 203

Met Ala Gly Gln Glu Asp Pro Val Gln Arg Glu Ile His Gln Asp Trp  
 1 5 10 15

Ala Asn Arg Glu Tyr Ile Glu Ile Ile Thr Ser Ser Ile Lys Lys Ile  
 20 25 30

Ala Asp Phe Leu Asn Ser Phe Asp Met Ser Cys Arg Ser Arg Leu Ala  
 35 40 45

Thr Leu Asn Glu Lys Leu Thr Ala Leu Glu Arg Arg Ile Glu Tyr Ile  
 50 55 60

Glu Ala Arg Val Thr Lys Gly Glu Thr Leu Thr  
 65 70 75

<210> 204

<211> 248

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (185)

<223> Xaa equals any of the naturally occurring L-amino acids

<400> 204

Met Thr Ser Gln Pro Val Pro Asn Glu Thr Ile Ile Val Leu Pro Ser  
 1 5 10 15

Asn Val Ile Asn Phe Ser Gln Ala Glu Lys Pro Glu Pro Thr Asn Gln  
 20 25 30

Gly Gln Asp Ser Leu Lys Lys His Leu His Ala Glu Ile Lys Val Ile  
 35 40 45

Gly Thr Ile Gln Ile Leu Cys Gly Met Met Val Leu Ser Leu Gly Ile  
 50 55 60

Ile Leu Ala Ser Ala Ser Phe Ser Pro Asn Phe Thr Gln Val Thr Ser

146

65                      70                      75                      80  
 Thr Leu Leu Asn Ser Ala Tyr Pro Phe Ile Gly Pro Phe Phe Phe Ile  
                                  85                      90                      95  
 Ile Ser Gly Ser Leu Ser Ile Ala Thr Glu Lys Arg Leu Thr Lys Leu  
                                  100                      105                      110  
 Leu Val His Ser Ser Leu Val Gly Ser Ile Leu Ser Ala Leu Ser Ala  
                                  115                      120                      125  
 Leu Val Gly Phe Ile Ile Leu Ser Val Lys Gln Ala Thr Leu Asn Pro  
                                  130                      135                      140  
 Ala Ser Leu Gln Cys Glu Leu Asp Lys Asn Asn Ile Pro Thr Arg Ser  
                                  145                      150                      155                      160  
 Tyr Val Ser Tyr Phe Tyr His Asp Ser Leu Tyr Thr Thr Asp Cys Tyr  
                                  165                      170                      175  
 Thr Ala Lys Ala Ser Leu Ala Gly Xaa Leu Ser Leu Met Leu Ile Cys  
                                  180                      185                      190  
 Thr Leu Leu Glu Phe Cys Leu Ala Val Leu Thr Ala Val Leu Arg Trp  
                                  195                      200                      205  
 Lys Gln Ala Tyr Ser Asp Phe Pro Gly Ser Val Leu Phe Leu Pro His  
                                  210                      215                      220  
 Ser Tyr Ile Gly Asn Ser Gly Met Ser Ser Lys Met Thr His Asp Cys  
                                  225                      230                      235                      240  
 Gly Tyr Glu Glu Leu Leu Thr Ser  
                                  245

&lt;210&gt; 205

&lt;211&gt; 168

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (83)

&lt;223&gt; Xaa equals any of the naturally occurring L-amino acids

&lt;400&gt; 205

Met Pro Leu Leu Arg Gly Leu Leu Trp Leu Gln Val Leu Cys Ala Gly  
   1                                 5                                 10                                 15  
 Pro Leu His Thr Glu Ala Val Val Leu Leu Val Pro Ser Asp Asp Gly  
                                  20                                 25                                 30  
 Arg Ala Phe Leu Leu Arg Ser Arg Leu Leu His Pro Glu Ala His Val  
                                  35                                 40                                 45  
 Pro Pro Ala Ala Asp Arg Gly Ala Ser Leu Gln Cys Val Leu His Gln  
                                  50                                 55                                 60



147

Ala Ala Pro Lys Ser Arg Pro Arg Ser Pro Ala Ala Gly Ala Ala Leu  
 65 70 75 80

Leu His Xaa Pro Arg Arg Thr Gly Asp Glu Pro Cys Arg Glu Phe His  
 85 90 95

Gly Asn Gly Phe Pro Gly Pro Thr Gln Leu Thr Pro Gly Glu Cys Gly  
 100 105 110

Leu Pro Ala Pro Ser Ser Leu Leu Gln His Ala Ser Ala Pro Val Arg  
 115 120 125

Thr Gly Ser Glu Gly Gln Val Val Gly Cys Pro Arg Ala Arg Gly Glu  
 130 135 140

Thr Gly Glu Gly Leu Ser Leu Ala Phe Leu Ser Ser Leu Met Phe Thr  
 145 150 155 160

Ser Arg Asn Gly Leu Val Gly Cys  
 165

<210> 206  
 <211> 218  
 <212> PRT  
 <213> Homo sapiens

<400> 206  
 Met Gly Ser Ala Ala Leu Glu Ile Leu Gly Leu Val Leu Cys Leu Val  
 1 5 10 15

Gly Trp Gly Gly Leu Ile Leu Ala Cys Gly Leu Pro Met Trp Gln Val  
 20 25 30

Thr Ala Phe Leu Asp His Asn Ile Val Thr Ala Gln Thr Thr Trp Lys  
 35 40 45

Gly Leu Trp Met Ser Cys Val Val Gln Ser Thr Gly His Met Gln Cys  
 50 55 60

Lys Val Tyr Asp Ser Val Leu Ala Leu Ser Thr Glu Val Gln Ala Ala  
 65 70 75 80

Arg Ala Leu Thr Val Ser Ala Val Leu Leu Ala Phe Val Ala Leu Phe  
 85 90 95

Val Thr Leu Ala Gly Ala Gln Cys Thr Thr Cys Val Ala Pro Gly Pro  
 100 105 110

Ala Lys Ala Arg Val Ala Leu Thr Gly Gly Val Leu Tyr Leu Phe Cys  
 115 120 125

Gly Leu Leu Ala Leu Val Pro Leu Cys Trp Phe Ala Asn Ile Val Val  
 130 135 140

Arg Glu Phe Tyr Asp Pro Ser Val Pro Val Ser Gln Lys Tyr Glu Leu  
 145 150 155 160

Gly Ala Ala Leu Tyr Ile Gly Trp Ala Ala Thr Ala Leu Leu Met Val

148

165 170 175

Gly Gly Cys Leu Leu Cys Cys Gly Ala Trp Val Cys Thr Gly Arg Pro  
180 185 190

Asp Leu Ser Phe Pro Val Lys Tyr Ser Ala Pro Arg Arg Pro Thr Ala  
195 200 205

Thr Gly Asp Tyr Asp Lys Lys Asn Tyr Val  
210 215

<210> 207  
<211> 73  
<212> PRT  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (73)  
<223> Xaa equals stop translation

<400> 207

Met Thr Ser Tyr Ile Leu Ile Ser Phe Val Leu Leu Ile Gly Val Gly  
1 5 10 15

Cys Ile Glu Lys Asp Gln Ser Cys Pro Val Phe Gly Gly Arg Lys Arg  
20 25 30

Leu His Leu Leu Phe Val Gly Gly Gln Leu Arg Gln Val Arg Met Leu  
35 40 45

Arg Gly Glu Leu Ser Cys Ala Cys Tyr Arg Pro His Val Gln Ala Leu  
50 55 60

Gln Leu Gly Gly Cys Thr Cys Phe Xaa  
65 70

<210> 208  
<211> 348  
<212> PRT  
<213> Homo sapiens

<400> 208

Met Leu Cys Pro Trp Arg Thr Ala Asn Leu Gly Leu Leu Leu Ile Leu  
1 5 10 15

Thr Ile Phe Leu Val Ala Glu Ala Glu Gly Ala Ala Gln Pro Asn Asn  
20 25 30

Ser Leu Met Leu Gln Thr Ser Lys Glu Asn His Ala Leu Ala Ser Ser  
35 40 45

Ser Leu Cys Met Asp Glu Lys Gln Ile Thr Gln Asn Tyr Ser Lys Val  
50 55 60

Leu Ala Glu Val Asn Thr Ser Trp Pro Val Lys Met Ala Thr Asn Ala  
65 70 75 80

Val Leu Cys Cys Pro Pro Ile Ala Leu Arg Asn Leu Ile Ile Ile Thr  
                                     85                                    90                                    95  
 Trp Glu Ile Ile Leu Arg Gly Gln Pro Ser Cys Thr Lys Ala Tyr Lys  
                                     100                                    105                                    110  
 Lys Glu Thr Asn Glu Thr Lys Glu Thr Asn Cys Thr Asp Glu Arg Ile  
                                     115                                    120                                    125  
 Thr Trp Val Ser Arg Pro Asp Gln Asn Ser Asp Leu Gln Ile Arg Thr  
                                     130                                    135                                    140  
 Val Ala Ile Thr His Asp Gly Tyr Tyr Arg Cys Ile Met Val Thr Pro  
                                     145                                    150                                    155                                    160  
 Asp Gly Asn Phe His Arg Gly Tyr His Leu Gln Val Leu Val Thr Pro  
                                     165                                    170                                    175  
 Glu Val Thr Leu Phe Gln Asn Arg Asn Arg Thr Ala Val Cys Lys Ala  
                                     180                                    185                                    190  
 Val Ala Gly Lys Pro Ala Ala His Ile Ser Trp Ile Pro Glu Gly Asp  
                                     195                                    200                                    205  
 Cys Ala Thr Lys Gln Glu Tyr Trp Ser Asn Gly Thr Val Thr Val Lys  
                                     210                                    215                                    220  
 Ser Thr Cys His Trp Glu Val His Asn Val Ser Thr Val Asn Cys His  
                                     225                                    230                                    235                                    240  
 Val Ser His Leu Thr Gly Asn Lys Ser Leu Tyr Ile Glu Leu Leu Pro  
                                     245                                    250                                    255  
 Val Pro Gly Ala Lys Lys Ser Ala Lys Leu Tyr Ile Pro Tyr Ile Ile  
                                     260                                    265                                    270

---

Leu Thr Ile Ile Ile Leu Thr Ile Val Gly Phe Ile Trp Leu Leu Lys  
                                     275                                    280                                    285  
 Val Asn Gly Cys Arg Lys Tyr Lys Leu Asn Lys Thr Glu Ser Thr Pro  
                                     290                                    295                                    300  
 Val Val Glu Glu Asp Glu Met Gln Pro Tyr Ala Ser Tyr Thr Glu Lys  
                                     305                                    310                                    315                                    320  
 Asn Asn Pro Leu Tyr Asp Thr Thr Asn Lys Val Lys Ala Ser Glu Ala  
                                     325                                    330                                    335  
 Leu Gln Ser Glu Val Asp Thr Asp Leu His Thr Leu  
                                     340                                    345

&lt;210&gt; 209

&lt;211&gt; 73

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

<221> SITE  
 <222> (73)  
 <223> Xaa equals stop translation

<400> 209

Met Ala Arg Gly Cys Val Cys Ser Leu Cys Ala Ser Val Cys Ile Phe  
 1 5 10 15

Leu Ser Ser Leu Phe Pro Leu Leu Pro Ser Val His Ser Val Asn Ile  
 20 25 30

Ile Ser Cys Leu Leu Leu Ser Lys Cys Phe Glu Gly Leu Glu Leu Met  
 35 40 45

Cys Glu His Leu Tyr Gln Leu Ser Gln Leu His Val Leu His His Ile  
 50 55 60

Phe Ser Tyr Leu Leu Cys Thr Pro Xaa  
 65 70

<210> 210  
 <211> 608  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (265)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (597)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<400> 210

Met Val Gly Thr Lys Leu Arg Gln Thr Lys Asp Ala Leu Phe Thr Ile  
 1 5 10 15

Leu His Asp Leu Arg Pro Gln Asp Arg Phe Ser Ile Ile Gly Phe Ser  
 20 25 30

Asn Arg Ile Lys Val Trp Lys Asp His Leu Ile Ser Val Thr Pro Asp  
 35 40 45

Ser Ile Arg Asp Gly Lys Val Tyr Ile His His Met Ser Pro Thr Gly  
 50 55 60

Gly Thr Asp Ile Asn Gly Val Leu Gln Arg Ala Ile Arg Leu Leu Asn  
 65 70 75 80

Lys Tyr Val Ala His Ser Gly Ile Gly Asp Arg Ser Val Ser Leu Ile  
 85 90 95

Val Phe Leu Thr Asp Gly Lys Pro Thr Val Gly Glu Thr His Thr Leu  
 100 105 110

Lys Ile Leu Asn Asn Thr Arg Glu Ala Ala Arg Gly Gln Val Cys Ile

115	120	125
Phe Thr Ile Gly Ile Gly Asn Asp Val Asp Phe Arg Leu Leu Glu Lys 130 135 140		
Leu Ser Leu Glu Asn Cys Gly Leu Thr Arg Arg Val His Glu Glu Glu 145 150 155 160		
Asp Ala Gly Ser Gln Leu Ile Gly Phe Tyr Asp Glu Ile Arg Thr Pro 165 170 175		
Leu Leu Ser Asp Ile Arg Ile Asp Tyr Pro Pro Ser Ser Val Val Gln 180 185 190		
Ala Thr Lys Thr Leu Phe Pro Asn Tyr Phe Asn Gly Ser Glu Ile Ile 195 200 205		
Ile Ala Gly Lys Leu Val Asp Arg Lys Leu Asp His Leu His Val Glu 210 215 220		
Val Thr Ala Ser Asn Ser Lys Lys Phe Ile Ile Leu Lys Thr Asp Val 225 230 235 240		
Pro Val Arg Pro Gln Lys Ala Gly Lys Asp Val Thr Gly Ser Pro Arg 245 250 255		
Pro Gly Gly Asp Gly Glu Gly Asp Xaa Asn His Ile Glu Arg Leu Trp 260 265 270		
Ser Tyr Leu Thr Thr Lys Glu Leu Leu Ser Ser Trp Leu Gln Ser Asp 275 280 285		
Asp Glu Pro Glu Lys Glu Arg Leu Arg Gln Arg Ala Gln Ala Leu Ala 290 295 300		
Val Ser Tyr Arg Phe Leu Thr Pro Phe Thr Ser Met Lys Leu Arg Gly 305 310 315 320		
Pro Val Pro Arg Met Asp Gly Leu Glu Glu Ala His Gly Met Ser Ala 325 330 335		
Ala Met Gly Pro Glu Pro Val Val Gln Ser Val Arg Gly Ala Gly Thr 340 345 350		
Gln Pro Gly Pro Leu Leu Lys Lys Pro Tyr Gln Pro Arg Ile Lys Ile 355 360 365		
Ser Lys Thr Ser Val Asp Gly Asp Pro His Phe Val Val Asp Phe Pro 370 375 380		
Leu Ser Arg Leu Thr Val Cys Phe Asn Ile Asp Gly Gln Pro Gly Asp 385 390 395 400		
Ile Leu Arg Leu Val Ser Asp His Arg Asp Ser Gly Val Thr Val Asn 405 410 415		
Gly Glu Leu Ile Gly Ala Pro Ala Pro Pro Asn Gly His Lys Lys Gln 420 425 430		

Arg Thr Tyr Leu Arg Thr Ile Thr Ile Leu Ile Asn Lys Pro Glu Arg  
 435 440 445  
 Ser Tyr Leu Glu Ile Thr Pro Ser Arg Val Ile Leu Asp Gly Gly Asp  
 450 455 460  
 Arg Leu Val Leu Pro Cys Asn Gln Ser Val Val Val Gly Ser Trp Gly  
 465 470 475 480  
 Leu Glu Val Ser Val Ser Ala Asn Ala Asn Val Thr Val Thr Ile Gln  
 485 490 495  
 Gly Ser Ile Ala Phe Val Ile Leu Ile His Leu Tyr Lys Lys Pro Ala  
 500 505 510  
 Pro Phe Gln Arg His His Leu Gly Phe Tyr Ile Ala Asn Ser Glu Gly  
 515 520 525  
 Leu Ser Ser Asn Cys His Gly Leu Leu Gly Gln Phe Leu Asn Gln Asp  
 530 535 540  
 Ala Arg Leu Thr Glu Asp Pro Ala Gly Pro Ser Gln Asn Leu Thr His  
 545 550 555 560  
 Pro Leu Leu Leu Gln Val Gly Glu Gly Pro Glu Ala Val Leu Thr Val  
 565 570 575  
 Lys Gly His Gln Val Pro Val Val Trp Lys Gln Arg Lys Ile Tyr Asn  
 580 585 590  
 Gly Glu Glu Gln Xaa Asp Cys Trp Phe Ala Arg Asn Met Pro Pro Asn  
 595 600 605

&lt;210&gt; 211

&lt;211&gt; 252

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (252)

&lt;223&gt; Xaa equals stop translation

&lt;400&gt; 211

Met Ala Pro Ala Ser Arg Leu Leu Ala Leu Trp Ala Leu Ala Ala Val  
 1 5 10 15  
 Ala Leu Pro Gly Ser Gly Ala Glu Gly Asp Gly Gly Trp Arg Pro Gly  
 20 25 30  
 Gly Pro Gly Ala Val Ala Glu Glu Glu Arg Cys Thr Val Glu Arg Arg  
 35 40 45  
 Ala Asp Leu Thr Tyr Ala Glu Phe Val Gln Gln Tyr Ala Phe Val Arg  
 50 55 60

Pro Val Ile Leu Gln Gly Leu Thr Asp Asn Ser Arg Phe Arg Ala Leu  
 65 70 75 80  
 Cys Ser Arg Asp Arg Leu Leu Ala Ser Phe Gly Asp Arg Val Val Arg  
 85 90 95  
 Leu Ser Thr Ala Asn Thr Tyr Ser Tyr His Lys Val Asp Leu Pro Phe  
 100 105 110  
 Gln Glu Tyr Val Glu Gln Leu Leu His Pro Gln Asp Pro Thr Ser Leu  
 115 120 125  
 Gly Asn Asp Thr Leu Tyr Phe Phe Gly Asp Asn Asn Phe Thr Glu Trp  
 130 135 140  
 Ala Ser Leu Phe Arg His Tyr Ser Pro Pro Pro Phe Gly Leu Leu Gly  
 145 150 155 160  
 Thr Ala Pro Ala Tyr Ser Phe Gly Ile Ala Gly Ala Gly Ser Gly Val  
 165 170 175  
 Pro Phe His Trp His Gly Pro Gly Tyr Ser Glu Val Ile Tyr Gly Arg  
 180 185 190  
 Lys Arg Trp Phe Leu Tyr Pro Pro Glu Lys Thr Pro Glu Phe His Pro  
 195 200 205  
 Asn Lys Thr Thr Leu Ala Trp Leu Arg Asp Thr Tyr Pro Ala Cys Thr  
 210 215 220  
 Val Cys Thr Ala Leu Glu Cys Thr Ile Arg Ala Gly Glu Val Leu Thr  
 225 230 235 240  
 Ser Arg Pro Leu Val Ala Cys Tyr Ala Gln Pro Xaa  
 245 250

<210> 212  
 <211> 226  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (226)  
 <223> Xaa equals stop translation

<400> 212  
 Met Lys Glu Ile Pro Ala Leu Leu His Leu Pro Val Leu Ile Ile Met  
 1 5 10 15  
 Ala Leu Ala Ile Leu Ser Phe Cys Tyr Gly Ala Gly Lys Ser Val His  
 20 25 30  
 Val Leu Arg His Ile Gly Gly Pro Glu Arg Glu Pro Pro Gln Ala Leu  
 35 40 45  
 Arg Pro Arg Asp Arg Arg Arg Gln Glu Glu Ile Asp Tyr Arg Pro Asp

```
<210> 213
<211> 51
<212> PRT
<213> Homo sapiens
```

```

<400> 213
Met Met Gly Leu Leu Glu Thr Gly Asn Val Leu Phe Trp Val Trp Val
 1              5              10              15
Val Val Thr Cys Val Tyr Ser Leu Tyr Ala Asn Ser Leu Asn Cys Thr
      20              25              30
Asp Met Asp Cys Ala Pro Phe Tyr Met Cys Val Met Leu Gln Gln Lys
      35              40              45
Cys Gln Xaa
      50

```



<210> 214  
 <211> 172  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (172)  
 <223> Xaa equals stop translation

<400> 214  
 Met Trp Leu Trp Ala Val Ser Pro Val Arg Pro Arg Thr Cys Leu Pro  
 1 5 10 15  
 Pro Cys Pro Arg Leu Trp Leu Trp Ile Ser Met Thr Leu Val Pro Ser  
 20 25 30  
 Ser Ser Ala Trp Lys Ser His Gly Ala Pro Ser Thr Arg Met Thr Ser  
 35 40 45  
 Pro Gln Leu Leu Leu Leu Ser Thr Arg Pro Pro Gln Ser Pro Ser Ala  
 50 55 60  
 Ser Pro Pro Ile Ala Arg Ala His Arg Thr His Pro His Phe Gly Asn  
 65 70 75 80  
 Arg Leu Ser Ile Thr Cys Cys Asp Gly Arg Arg Ser Trp Arg Met Gly  
 85 90 95  
 Gln His Gly Pro Cys His Leu Asn Leu Gln Thr Thr His Pro Ala His  
 100 105 110  
 Ser Ser Gln Ala Leu Pro Ala Thr His Gln Pro Leu Gly Pro Trp Cys  
 115 120 125

---

Ser Ser Pro Ser Pro Phe Pro Ser Lys Leu Pro Ser Ala Gly Leu Arg  
 130 135 140  
 Pro Pro Ala Leu Gly Pro Trp Met Arg Arg Gly Pro Trp Pro Gln Ser  
 145 150 155 160  
 Trp Gln Met Gly Met His Pro Thr Val Gly Leu Xaa  
 165 170

<210> 215  
 <211> 48  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (48)  
 <223> Xaa equals stop translation

<400> 215  
 Met Trp Leu Leu Ile Ile Phe Cys Lys Ser Ala Ser Ala Ser Val Leu  
 1 5 10 15

Cys Trp Ile Lys Lys Phe His Pro Val Phe Gln Glu Ser Leu Leu Tyr  
                   20                                  25                                  30

Leu Val Gln Glu Gly Ser Leu Cys Tyr Val Gln Gln Lys Val Pro Xaa  
                   35                                  40                                  45

<210> 216  
 <211> 139  
 <212> PRT  
 <213> Homo sapiens

<400> 216  
 Met Glu Ala Val Val Phe Val Phe Ser Leu Leu Asp Cys Cys Ala Leu  
   1                                  5                                  10                                  15  
 Ile Phe Leu Ser Val Tyr Phe Ile Ile Thr Leu Ser Asp Leu Glu Cys  
                   20                                  25                                  30  
 Asp Tyr Ile Asn Ala Arg Ser Cys Cys Ser Lys Leu Asn Lys Trp Val  
                   35                                  40                                  45  
 Ile Pro Glu Leu Ile Gly His Thr Ile Val Thr Val Leu Leu Leu Met  
                   50                                  55                                  60  
 Ser Leu His Trp Phe Ile Phe Leu Leu Asn Leu Pro Val Ala Thr Trp  
   65                                  70                                  75                                  80  
 Asn Ile Tyr Arg Tyr Ile Met Val Pro Ser Gly Asn Met Gly Val Phe  
                   85                                  90                                  95  
 Asp Pro Thr Glu Ile His Asn Arg Gly Gln Leu Lys Ser His Met Lys  
                   100                                  105                                  110  
 Glu Ala Met Ile Lys Leu Gly Phe His Leu Leu Cys Phe Phe Met Tyr  
                   115                                  120                                  125  
 Leu Tyr Ser Met Ile Leu Ala Leu Ile Asn Asp  
                   130                                  135

<210> 217  
 <211> 41  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (41)  
 <223> Xaa equals stop translation

<400> 217  
 Met Ser Gly Ser Ser Leu Pro Ser Ala Leu Ala Leu Ser Leu Leu Leu  
   1                                  5                                  10                                  15

Val Ser Gly Ser Leu Leu Pro Gly Pro Gly Ala Ala Gln Asn Val Arg  
20 25 30

Val Gln Ser Gly Gln Asp Gln Lys Xaa  
35 40

<210> 218  
<211> 52  
<212> PRT  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (52)  
<223> Xaa equals stop translation

<400> 218  
Met Pro Ser His Ile Arg Ala His Leu Phe Leu Leu Leu Phe Phe Leu  
1 5 10 15

Phe Ile Tyr Gln Gly Ile Ser Ser Ile Ser Gln Ala Ser Gly Leu Thr  
20 25 30

Leu Lys Thr Gln Asn Glu Lys Asp Ile Gln Val Ser Ile Leu Lys Glu  
35 40 45

Phe Val Val Xaa  
50

<210> 219  
<211> 49  
<212> PRT  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (49)  
<223> Xaa equals stop translation

<400> 219  
Met Cys Ile Tyr Gln Ser Glu Gln Met Leu Ala Leu Leu Leu Val Leu  
1 5 10 15

Val Phe Cys Ile Ser Leu Leu Val Leu Val Cys Trp Gly Ser His Asn  
20 25 30

Lys Val Pro Gln Lys Phe Ile Phe Ser Gln Phe Trp Gly Leu Glu Asp  
35 40 45

Xaa

<210> 220  
<211> 42  
<212> PRT  
<213> Homo sapiens

<220>  
 <221> SITE  
 <222> (42)  
 <223> Xaa equals stop translation

<400> 220  
 Met Ala Val Pro Leu Phe Leu Tyr Ile Phe Thr Leu Leu Pro Leu Leu  
 1 5 10 15  
 Pro Phe Leu Leu Ser Leu Cys Phe Ser Pro Leu Thr Val Lys Arg Ser  
 20 25 30  
 Ser Ser Ser Glu Ser Lys Ser Ser Leu Xaa  
 35 40

<210> 221  
 <211> 41  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (41)  
 <223> Xaa equals stop translation

<400> 221  
 Met Gly Met Leu Leu Ala Phe Trp Leu Pro Gly Ala Ser Trp Gln Glu  
 1 5 10 15  
 Ala Gly Pro Arg Ala Ser Thr Gln Arg Met Arg Thr Gln Thr Gln Met  
 20 25 30  
 Ser Thr Arg Lys Pro Lys Pro Ala Xaa  
 35 40

<210> 222  
 <211> 43  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (43)  
 <223> Xaa equals stop translation

<400> 222  
 Met Glu Pro Ala Met Val Leu Lys Phe Leu Ser Ser Leu Pro Glu Asn  
 1 5 10 15  
 Leu Phe Leu Pro Ser Leu Leu Phe Phe Ala Trp Leu Cys Trp Asn Met  
 20 25 30  
 Val Cys Gly Ser Pro Val Ser Cys Pro Tyr Xaa  
 35 40

<210> 223  
 <211> 204  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (204)  
 <223> Xaa equals stop translation

<400> 223  
 Met Gln Leu Gly Ser Val Leu Leu Thr Arg Cys Pro Phe Trp Gly Cys  
   1                  5                  10                  15  
 Phe Ser Gln Leu Met Leu Tyr Ala Glu Arg Ala Glu Ala Arg Arg Lys  
           20                  25                  30  
 Pro Asp Ile Pro Val Pro Tyr Leu Tyr Phe Asp Met Gly Ala Ala Val  
           35                  40                  45  
 Leu Cys Ala Ser Phe Met Ser Phe Gly Val Lys Arg Arg Trp Phe Ala  
   50                  55                  60  
 Leu Gly Ala Ala Leu Gln Leu Ala Ile Ser Thr Tyr Ala Ala Tyr Ile  
   65                  70                  75                  80  
 Gly Gly Tyr Val His Tyr Gly Asp Trp Leu Lys Val Arg Met Tyr Ser  
                   85                  90                  95  
 Arg Thr Val Ala Ile Ile Gly Gly Phe Leu Val Leu Ala Ser Gly Ala  
           100                  105                  110  
 Gly Glu Leu Tyr Arg Arg Lys Pro Arg Ser Arg Ser Leu Gln Ser Thr  
   115                  120                  125  
 Gly Gln Val Phe Leu Gly Ile Tyr Leu Ile Cys Val Ala Tyr Ser Leu  
   130                  135                  140

---

Gln His Ser Lys Glu Asp Arg Leu Ala Tyr Leu Asn His Leu Pro Gly  
 145                  150                  155                  160  
 Gly Glu Leu Met Ile Gln Leu Phe Phe Val Leu Tyr Gly Ile Leu Ala  
           165                  170                  175  
 Pro Gly Leu Ser Val Arg Leu Leu Arg Asp Pro Arg Cys Pro Asp Pro  
   180                  185                  190  
 Gly Cys Thr Ala Ala Pro Cys His Ala Ala His Xaa  
   195                  200

<210> 224  
 <211> 43  
 <212> PRT  
 <213> Homo sapiens.

<220>  
 <221> SITE  
 <222> (43)

<223> Xaa equals stop translation

<400> 224

Met Arg Val Arg Ile Gly Leu Thr Leu Leu Leu Cys Ala Val Leu Leu  
1 5 10 15

Ser Leu Ala Ser Ala Ser Ser Asp Glu Glu Gly Ser Gln Asp Glu Ser  
20 25 30

Leu Gly Phe Gln Asp Tyr Phe Asp Ile Arg Xaa  
35 40

<210> 225

<211> 156

<212> PRT

<213> Homo sapiens

<220>

<221> SITE

<222> (156)

<223> Xaa equals stop translation

<400> 225

Met Ala Arg Gly Ser Leu Arg Arg Leu Leu Arg Leu Leu Val Leu Gly  
1 5 10 15

Leu Trp Leu Ala Leu Leu Arg Ser Val Ala Gly Glu Gln Ala Pro Gly  
20 25 30

Thr Ala Pro Cys Ser Arg Gly Ser Ser Trp Ser Ala Asp Leu Asp Lys  
35 40 45

Cys Met Asp Cys Ser Thr Ser Cys Pro Leu Pro Ala Ala Leu Ala His  
50 55 60

Pro Trp Gly Arg Ser Glu Pro Asp Leu Arg Ala Gly Ala Ala Phe Trp  
65 70 75 80

Leu Phe Gly Leu Glu Thr Met Pro Gln Glu Arg Glu Val His His Pro  
85 90 95

His Arg Gly Asp Arg Arg Arg Gly Leu Pro Ser Cys Gly Ala Asp Pro  
100 105 110

Val Thr Met Cys Pro Leu Pro Ala Gly Ala Arg Pro Leu Ile Ile His  
115 120 125

Ser Ser Ile Leu Glu Pro Val Ser Ala Ser Gln Thr Arg Arg Glu Pro  
130 135 140

Ser Ser Ser Asn His Lys Gly Gly Gly Gly Arg Xaa  
145 150 155

<210> 226

<211> 74

<212> PRT

<213> Homo sapiens

<220>  
 <221> SITE  
 <222> (38)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (48)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (54)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (55)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (68)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (74)  
 <223> Xaa equals stop translation

<400> 226  
 Met Phe Tyr Lys Leu Thr Leu Ile Leu Cys Glu Leu Ser Val Ala Gly  
           1                  5                  10                  15

Val Thr Gln Ala Ala Ser Gln Arg Pro Leu Gln Arg Leu Pro Arg His  
                   20                  25                  30

Ile Cys Ser Gln Arg Xaa Pro Pro Gly Arg Cys Leu Leu Lys Ala Xaa  
           35                  40                  45

Leu Gln Thr Thr Trp Xaa Xaa Pro Asp Lys Pro Ile Pro Arg Leu Ser  
           50                  55                  60

Pro Pro Leu Xaa Ser Asp Pro Lys Arg Xaa  
           65                  70

<210> 227  
 <211> 167  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (167)  
 <223> Xaa equals stop translation

<400> 227

162

Met Gly Ser Arg Phe Leu Leu Val Leu Leu Ser Gly Leu Thr Val Leu  
 1 5 10 15  
 Leu Ala Leu Pro Gly Ser Glu Ala Lys Asn Ser Gly Ala Ser Cys Pro  
 20 25 30  
 Pro Cys Pro Lys Tyr Ala Ser Cys His Asn Ser Thr His Cys Thr Cys  
 35 40 45  
 Glu Asp Gly Phe Arg Ala Arg Ser Gly Arg Thr Tyr Phe His Asp Ser  
 50 55 60  
 Ser Glu Lys Cys Glu Asp Ile Asn Glu Cys Glu Thr Gly Leu Ala Lys  
 65 70 75 80  
 Cys Lys Tyr Lys Ala Tyr Cys Arg Asn Lys Val Gly Gly Tyr Ile Cys  
 85 90 95  
 Ser Cys Leu Val Lys Tyr Thr Leu Phe Asn Phe Leu Ala Gly Ile Ile  
 100 105 110  
 Asp Tyr Asp His Pro Asp Cys Tyr Glu Asn Asn Ser Gln Gly Thr Thr  
 115 120 125  
 Gln Ser Asn Val Asp Ile Trp Val Ser Gly Val Lys Pro Gly Phe Gly  
 130 135 140  
 Lys Gln Leu Val Arg Ile Thr Met Pro Phe Ser Tyr Pro Asn Ile Asn  
 145 150 155 160  
 Met Ser Ser Cys Asp Phe Xaa  
 165

&lt;210&gt; 228

&lt;211&gt; 71

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (71)

&lt;223&gt; Xaa equals stop translation

&lt;400&gt; 228

Met Lys Pro Lys His Leu Glu Trp Cys Leu Ala His Ser Trp Cys Val  
 1 5 10 15  
 Ile Trp Leu Ser Phe Val Ser Pro Pro Thr Ser His Leu Glu Cys Asp  
 20 25 30  
 Gly Phe Pro Gly Ser Leu Leu Pro Pro Cys Glu Glu Gly Arg Cys Phe  
 35 40 45  
 Pro Phe Thr Phe His His His Asp Cys His Gly Cys Ser Pro Leu Gln  
 50 55 60  
 Ser Ser Pro Gly Gln His Xaa  
 65 70



<210> 229  
 <211> 273  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (273)  
 <223> Xaa equals stop translation

<400> 229

Met Cys Cys Trp Pro Leu Leu Leu Leu Trp Gly Leu Leu Pro Gly Thr  
 1 5 10 15

Ala Ala Gly Gly Ser Gly Arg Thr Tyr Pro His Arg Thr Leu Leu Asp  
 20 25 30

Ser Glu Gly Lys Tyr Trp Leu Gly Trp Ser Gln Arg Gly Ser Gln Ile  
 35 40 45

Ala Phe Arg Leu Gln Val Arg Thr Ala Gly Tyr Val Gly Phe Gly Phe  
 50 55 60

Ser Pro Thr Gly Ala Met Ala Ser Ala Asp Ile Val Val Gly Gly Val  
 65 70 75 80

Ala His Gly Arg Pro Tyr Leu Gln Asp Tyr Phe Thr Asn Ala Asn Arg  
 85 90 95

Glu Leu Lys Lys Asp Ala Gln Gln Asp Tyr His Leu Glu Tyr Ala Met  
 100 105 110

Glu Asn Ser Thr His Thr Ile Ile Glu Phe Thr Arg Glu Leu His Thr  
 115 120 125

Cys Asp Ile Asn Asp Lys Ser Ile Thr Asp Ser Thr Val Arg Val Ile  
 130 135 140

Trp Ala Tyr His His Glu Asp Ala Gly Glu Ala Gly Pro Lys Tyr His  
 145 150 155 160

Asp Ser Asn Arg Gly Thr Lys Ser Leu Arg Leu Leu Asn Pro Glu Lys  
 165 170 175

Thr Ser Val Leu Ser Thr Ala Leu Pro Tyr Phe Asp Leu Val Asn Gln  
 180 185 190

Asp Val Pro Ile Pro Asn Lys Asp Thr Thr Tyr Trp Cys Gln Met Phe  
 195 200 205

Lys Ile Pro Val Phe Gln Glu Lys His His Val Ile Lys Val Glu Pro  
 210 215 220

Val Ile Gln Arg Gly His Glu Ser Leu Val His His Ile Leu Leu Tyr  
 225 230 235 240

Gln Cys Ser Asn Asn Phe Asn Asp Ser Val Pro Gly Ile Arg Ala Arg

164

245 250 255  
 Ile Ala Ile Thr Pro Thr Cys Pro Met His Ser Ser Pro Val Lys Leu  
 260 265 270

Xaa

<210> 230  
 <211> 82  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (82)  
 <223> Xaa equals stop translation

<400> 230  
 Met Arg Pro Gly Thr Ala Leu Gln Ala Val Leu Leu Ala Val Leu Leu  
 1 5 10 15

Val Gly Leu Arg Ala Ala Thr Gly Arg Leu Leu Ser Gly Gln Pro Val  
 20 25 30

Cys Arg Gly Gly Thr Gln Arg Pro Cys Tyr Lys Val Ile Tyr Phe His  
 35 40 45

Asp Thr Ser Arg Arg Leu Asn Phe Glu Glu Ala Lys Glu Ala Cys Arg  
 50 55 60

Arg Gly Trp Arg Pro Ala Ser Gln His Arg Val Leu Lys Met Asn Arg  
 65 70 75 80

Asn Xaa

<210> 231  
 <211> 71  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (38)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<400> 231  
 Met Ser Pro Leu Ser Ala Ala Arg Ala Ala Leu Arg Val Tyr Ala Val  
 1 5 10 15

Gly Ala Ala Val Ile Leu Ala Gln Leu Leu Arg Arg Cys Arg Gly Gly  
 20 25 30

Phe Leu Glu Pro Val Xaa Pro Pro Arg Pro Asp Arg Val Ala Ile Val  
 35 40 45

165

Thr Gly Gly Thr Asp Gly Ile Gly Tyr Ser Thr Ala Asn Ile Trp Arg  
 50 55 60

Asp Leu Ala Cys Met Leu Ser  
 65 70

&lt;210&gt; 232

&lt;211&gt; 225

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; SITE

&lt;222&gt; (5)

&lt;223&gt; Xaa equals any of the naturally occurring L-amino acids

&lt;400&gt; 232

His Glu Arg Ala Xaa Gly Pro Ser Arg Gly His Gly Glu Leu Leu Ser  
 1 5 10 15

Cys Val Leu Gly Pro Arg Leu Tyr Lys Ile Tyr Arg Glu Arg Asp Ser  
 20 25 30

Glu Arg Ala Pro Ala Ser Val Pro Glu Thr Pro Thr Ala Val Thr Ala  
 35 40 45

Pro His Ser Ser Ser Trp Asp Thr Tyr Tyr Gln Pro Arg Ala Leu Glu  
 50 55 60

Lys His Ala Asp Ser Ile Leu Ala Leu Ala Ser Val Phe Trp Ser Ile  
 65 70 75 80

Ser Tyr Tyr Ser Ser Pro Phe Ala Phe Phe Tyr Leu Tyr Arg Lys Gly  
 85 90 95

Tyr Leu Ser Leu Ser Lys Val Val Pro Phe Ser His Tyr Ala Gly Thr  
 100 105 110

Leu Leu Leu Leu Leu Ala Gly Val Ala Cys Ser Glu Ala Leu Ala Ala  
 115 120 125

Gly Pro Thr Pro Ser Thr Gly Ser Ser Ser Pro Ser Trp Lys Gln His  
 130 135 140

Ile Gly Thr Ser Leu Gln Lys Thr Arg Gly Ser Leu Pro Thr Thr Thr  
 145 150 155 160

Leu Thr Ser Gly Ala Gly Gln Ser Thr Ser Thr Gly Lys Asn Pro Ala  
 165 170 175

Ala Gly Arg Ser Leu Glu Gly Ala Leu Pro Ala Gly Val Trp Pro Cys  
 180 185 190

Phe Ala Gln Ser Pro Cys Thr Gly Gly Gln Gln Thr Pro Ser Ser Thr  
 195 200 205

Gly Leu Arg Ser Cys Leu Val Arg Ser Pro Ala Thr Trp Trp Arg Thr  
 210 215 220

Pro  
225

<210> 233  
<211> 314  
<212> PRT  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (147)  
<223> Xaa equals any of the naturally occurring L-amino acids

<220>  
<221> SITE  
<222> (211)  
<223> Xaa equals any of the naturally occurring L-amino acids

<400> 233

Met Leu Pro Ala Arg Leu Pro Phe Arg Leu Leu Ser Leu Phe Leu Arg  
1 5 10 15

Gly Ser Ala Pro Thr Ala Ala Arg His Gly Leu Arg Glu Pro Leu Leu  
20 25 30

Glu Arg Arg Cys Ala Ala Ala Ser Ser Phe Gln His Ser Ser Ser Leu  
35 40 45

Gly Arg Glu Leu Pro Tyr Asp Pro Val Asp Thr Glu Gly Phe Gly Glu  
50 55 60

Gly Gly Asp Met Gln Glu Arg Phe Leu Phe Pro Glu Tyr Ile Leu Asp  
65 70 75 80

Pro Glu Pro Gln Pro Thr Arg Glu Lys Gln Leu Gln Glu Leu Gln Gln  
85 90 95

Gln Gln Glu Glu Glu Glu Arg Gln Arg Gln Gln Arg Arg Glu Glu Arg  
100 105 110

Arg Gln Gln Asn Leu Arg Ala Arg Ser Arg Glu His Pro Val Val Gly  
115 120 125

His Pro Asp Pro Ala Leu Pro Pro Ser Gly Val Asn Cys Ser Gly Cys  
130 135 140

Gly Ala Xaa Leu His Cys Gln Asp Ala Gly Val Pro Gly Tyr Leu Pro  
145 150 155 160

Arg Glu Lys Phe Leu Arg Thr Ala Glu Ala Asp Gly Gly Leu Ala Arg  
165 170 175

Thr Val Cys Gln Arg Cys Trp Leu Leu Ser His His Arg Arg Ala Leu  
180 185 190

Arg Leu Gln Val Ser Arg Glu Gln Tyr Leu Glu Leu Val Ser Ala Ala  
195 200 205

Leu Arg Xaa Pro Gly Pro Ser Leu Val Leu Tyr Met Val Asp Leu Leu  
 210 215 220  
 Asp Leu Pro Asp Ala Leu Leu Pro Asp Leu Pro Ala Leu Val Gly Pro  
 225 230 235 240  
 Lys Gln Leu Ile Val Leu Gly Asn Lys Val Asp Leu Leu Pro Gln Asp  
 245 250 255  
 Ala Pro Gly Tyr Arg Gln Arg Leu Arg Glu Arg Leu Trp Glu Asp Cys  
 260 265 270  
 Ala Arg Ala Gly Leu Leu Leu Ala Pro Gly Thr Lys Gly His Ser Ala  
 275 280 285  
 Pro Ser Arg Thr Ser His Arg Thr Gly Arg Ile Arg Ile Arg Arg Thr  
 290 295 300  
 Gly Pro Ala Gln Trp Ser Gly Thr Cys Gly  
 305 310

<210> 234  
 <211> 93  
 <212> PRT  
 <213> Homo sapiens

<400> 234  
 Met Arg Pro Gln Gly Pro Ala Ala Ser Pro Gln Arg Leu Arg Gly Leu  
 1 5 10 15  
 Leu Leu Leu Leu Leu Leu Gln Leu Pro Ala Pro Ser Ser Ala Ser Glu  
 20 25 30  
 Ile Pro Lys Gly Lys Gln Lys Ala His Ser Gly Arg Gly Arg Trp Trp  
 35 40 45  
 Thr Cys Ile Met Glu Cys Ala Tyr Lys Gly Gln Gln Glu Cys Leu Val  
 50 55 60  
 Glu Thr Gly Ala Leu Gly Pro Met Ala Phe Arg Val His Leu Gly Ser  
 65 70 75 80  
 Gln Val Gly Met Asp Ser Lys Glu Lys Arg Gly Asn Val  
 85 90

<210> 235  
 <211> 73  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (73)  
 <223> Xaa equals stop translation  
 <400> 235

Met Gly Ser Ala Ala Leu Glu Ile Leu Gly Leu Val Leu Cys Leu Val  
 1 5 10 15  
 Gly Trp Gly Gly Leu Ile Leu Ala Cys Gly Leu Pro Met Trp Gln Val  
 20 25 30  
 Thr Ala Phe Leu Asp His Asn Ile Val Thr Ala Gln Thr Thr Trp Lys  
 35 40 45  
 Gly Leu Trp Met Ser Cys Val Val Gln Ser Thr Gly Thr Cys Ser Ala  
 50 55 60  
 Lys Cys Thr Thr Arg Cys Trp Leu Xaa  
 65 70

<210> 236  
 <211> 349  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (283)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (293)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (325)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (326)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (349)  
 <223> Xaa equals stop translation

<400> 236  
 Met Leu Cys Pro Trp Arg Thr Ala Asn Leu Gly Leu Leu Leu Ile Leu  
 1 5 10 15  
 Thr Ile Phe Leu Val Ala Glu Ala Glu Gly Ala Ala Gln Pro Asn Asn  
 20 25 30  
 Ser Leu Met Leu Gln Thr Ser Lys Glu Asn His Ala Leu Ala Ser Ser  
 35 40 45  
 Ser Leu Cys Met Asp Glu Lys Gln Ile Thr Gln Asn Tyr Ser Lys Val  
 50 55 60

Leu Ala Glu Val Asn Thr Ser Trp Pro Val Lys Met Ala Thr Asn Ala  
 65 70 75 80  
 Val Leu Cys Cys Pro Pro Ile Ala Leu Arg Asn Leu Ile Ile Ile Thr  
 85 90 95  
 Trp Glu Ile Ile Leu Arg Gly Gln Pro Ser Cys Thr Lys Ala Tyr Lys  
 100 105 110  
 Lys Glu Thr Asn Glu Thr Lys Glu Thr Asn Cys Thr Asp Glu Arg Ile  
 115 120 125  
 Thr Trp Val Ser Arg Pro Asp Gln Asn Ser Asp Leu Gln Ile Arg Thr  
 130 135 140  
 Val Ala Ile Thr His Asp Gly Tyr Tyr Arg Cys Ile Met Val Thr Pro  
 145 150 155 160  
 Asp Gly Asn Phe His Arg Gly Tyr His Leu Gln Val Leu Val Thr Pro  
 165 170 175  
 Glu Val Thr Leu Phe Gln Asn Arg Asn Arg Thr Ala Val Cys Lys Ala  
 180 185 190  
 Val Ala Gly Lys Pro Ala Ala His Ile Ser Trp Ile Pro Glu Gly Asp  
 195 200 205  
 Cys Ala Thr Lys Gln Glu Tyr Trp Ser Asn Gly Thr Val Thr Val Lys  
 210 215 220  
 Ser Thr Cys His Trp Glu Val His Asn Val Ser Thr Val Asn Cys His  
 225 230 235 240  
 Val Ser His Leu Thr Gly Asn Lys Ser Leu Tyr Ile Glu Leu Leu Pro  
 245 250 255  


---

 Val Pro Gly Ala Lys Lys Ser Ser Lys Leu Tyr Ile Pro Tyr Ile Ile  
 260 265 270  
 Leu Thr Ile Ile Ile Leu Thr Ile Val Gly Xaa Ile Trp Leu Leu Lys  
 275 280 285  
 Val Asn Gly Cys Xaa Lys Tyr Lys Leu Asn Lys Pro Glu Ser Thr Pro  
 290 295 300  
 Val Val Glu Glu Asp Glu Met Gln Pro Tyr Ala Phe Tyr Thr Glu Lys  
 305 310 315 320  
 Asn Asn Pro Leu Xaa Xaa Thr Thr Asn Lys Val Lys Ala Ser Glu Ala  
 325 330 335  
 Leu Gln Ser Glu Val Asp Thr Asp Leu His Thr Leu Xaa  
 340 345

&lt;210&gt; 237

&lt;211&gt; 17

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 237

Leu Ala Leu Tyr Ser Ala Leu Phe Ser Tyr Ser Gly Trp Asp Thr Leu  
1 5 10 15

Asn

&lt;210&gt; 238

&lt;211&gt; 14

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 238

Val Thr Glu Glu Ile Lys Asn Pro Glu Arg Asn Leu Pro Leu  
1 5 10

&lt;210&gt; 239

&lt;211&gt; 9

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 239

Ile Gly Ile Ser Met Pro Ile Val Thr  
1 5

&lt;210&gt; 240

&lt;211&gt; 13

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 240

Ile Tyr Ile Leu Thr Asn Val Ala Tyr Tyr Thr Val Leu  
1 5 10

&lt;210&gt; 241

&lt;211&gt; 11

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 241

Ser Asp Ala Val Ala Val Thr Phe Ala Asp Gln  
1 5 10

&lt;210&gt; 242

&lt;211&gt; 13

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 242

Val Ala Leu Ser Cys Phe Gly Gly Leu Asn Ala Ser Ile  
1 5 10



<210> 243  
<211> 15  
<212> PRT  
<213> Homo sapiens

<400> 243  
Ser Arg Leu Phe Phe Val Gly Ser Arg Glu Gly His Leu Pro Asp  
1 5 10 15

<210> 244  
<211> 11  
<212> PRT  
<213> Homo sapiens

<400> 244  
Ser Phe Ser Tyr Trp Phe Phe Val Gly Leu Ser  
1 5 10

<210> 245  
<211> 11  
<212> PRT  
<213> Homo sapiens

<400> 245  
Val Gly Gln Leu Tyr Leu Arg Trp Lys Glu Pro  
1 5 10

<210> 246  
<211> 16  
<212> PRT  
<213> Homo sapiens

<400> 246  
Arg Pro Arg Pro Leu Lys Leu Ser Val Phe Phe Pro Ile Val Phe Cys  
1 5 10 15

<210> 247  
<211> 9  
<212> PRT  
<213> Homo sapiens

<400> 247  
Asp Thr Ile Asn Ser Leu Ile Gly Ile  
1 5

<210> 248  
<211> 44  
<212> PRT  
<213> Homo sapiens

<400> 248

Ala Thr Ala Leu Pro Pro Lys Ile Val Gly Ser Ala Thr Arg Tyr Leu  
1 5 10 15

Gln Val Leu Cys Met Ser Val Ala Ala Glu Met Asp Leu Glu Asp Gly  
20 25 30

Gly Glu Met Pro Lys Gln Arg Asp Pro Lys Ser Asn  
35 40

<210> 249

<211> 352

<212> PRT

<213> Homo sapiens

<400> 249

Leu Leu Ala Ala Ala Cys Ile Cys Leu Leu Thr Phe Ile Asn Cys Ala  
1 5 10 15

Tyr Val Lys Trp Gly Thr Leu Val Gln Asp Ile Phe Thr Tyr Ala Lys  
20 25 30

Val Leu Ala Leu Ile Ala Val Ile Val Ala Gly Ile Val Arg Leu Gly  
35 40 45

Gln Gly Ala Ser Thr His Phe Glu Asn Ser Phe Glu Gly Ser Ser Phe  
50 55 60

Ala Val Gly Asp Ile Ala Leu Ala Leu Tyr Ser Ala Leu Phe Ser Tyr  
65 70 75 80

Ser Gly Trp Asp Thr Leu Asn Tyr Val Thr Glu Glu Ile Lys Asn Pro  
85 90 95

Glu Arg Asn Leu Pro Leu Ser Ile Gly Ile Ser Met Pro Ile Val Thr  
100 105 110

Ile Ile Tyr Ile Leu Thr Asn Val Ala Tyr Tyr Thr Val Leu Asp Met  
115 120 125

Arg Asp Ile Leu Ala Ser Asp Ala Val Ala Val Thr Phe Ala Asp Gln  
130 135 140

Ile Phe Gly Ile Phe Asn Trp Ile Ile Pro Leu Ser Val Ala Leu Ser  
145 150 155 160

Cys Phe Gly Gly Leu Asn Ala Ser Ile Val Ala Ala Ser Arg Leu Phe  
165 170 175

Phe Val Gly Ser Arg Glu Gly His Leu Pro Asp Ala Ile Cys Met Ile  
180 185 190

His Val Glu Arg Phe Thr Pro Val Pro Ser Leu Leu Phe Asn Gly Ile  
195 200 205

Met Ala Leu Ile Tyr Leu Cys Val Glu Asp Ile Phe Gln Leu Ile Asn  
210 215 220

Tyr Tyr Ser Phe Ser Tyr Trp Phe Phe Val Gly Leu Ser Ile Val Gly

```
<210> 250
<211> 119
<212> PRT
<213> Homo sapiens
```

<400> 250  
Ala Ala Arg Gly Ser Gly Val Arg Asp Pro Leu Glu Glu Ala Val Cys  
1 5 10 15

[illegible]

<210> 251

<211> 356  
 <212> PRT  
 <213> Homo sapiens

<220>  
 <221> SITE  
 <222> (37)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<220>  
 <221> SITE  
 <222> (280)  
 <223> Xaa equals any of the naturally occurring L-amino acids

<400> 251

```

Leu Ser Lys Ala Phe Leu Asp Ser Pro Asn Arg Leu Leu Ala Val Glu
 1             5             10             15

Met Asn Thr Asp His Leu Arg Leu Thr Val Pro Asn Gly Ile Gly Ala
      20             25             30

Leu Lys Leu Arg Xaa Met Glu His Tyr Phe Ser Gln Gly Leu Ser Val
      35             40             45

Gln Leu Phe Asn Asp Gly Ser Lys Gly Lys Leu Asn His Leu Cys Gly
      50             55             60

Ala Asp Phe Val Lys Ser His Gln Lys Pro Pro Gln Gly Met Glu Ile
      65             70             75             80

Lys Ser Asn Glu Arg Cys Cys Ser Phe Asp Gly Asp Ala Asp Arg Ile
      85             90             95

Val Tyr Tyr Tyr His Asp Ala Asp Gly His Phe His Leu Ile Asp Gly
      100            105            110

Asp Lys Ile Ala Thr Leu Ile Ser Ser Phe Leu Lys Glu Leu Leu Val
      115            120            125

Glu Ile Gly Glu Ser Leu Asn Ile Gly Val Val Gln Thr Ala Tyr Ala
      130            135            140

Asn Gly Ser Ser Thr Arg Tyr Leu Glu Glu Val Met Lys Val Pro Val
      145            150            155            160

Tyr Cys Thr Lys Thr Gly Val Lys His Leu His His Lys Ala Gln Glu
      165            170            175

Phe Asp Ile Gly Val Tyr Phe Glu Ala Asn Gly His Gly Thr Ala Leu
      180            185            190

Phe Ser Thr Ala Val Glu Met Lys Ile Lys Gln Ser Ala Glu Gln Leu
      195            200            205

Glu Asp Lys Lys Arg Lys Ala Ala Lys Met Leu Glu Asn Ile Ile Asp
      210            215            220

Leu Phe Asn Gln Ala Ala Gly Asp Ala Ile Ser Asp Met Leu Val Ile
      225            230            235            240

```

[illegible]

```
<210> 252
<211> 26
<212> PRT
<213> Homo sapiens
```

```

<400> 252
Leu Ser Lys Ala Phe Leu Asp Ser Pro Asn Arg Leu Leu Ala Val Glu
 1             5             10             15

```

Met Asn Thr Asp His Leu Arg Leu Thr Val  
20 25

```
<210> 253
<211> 28
<212> PRT
<213> Homo sapiens
```

```
<220>
<221> SITE
<222> (11)
<223> Xaa equals any of the naturally occurring L-amino acids
```

```
<400> 253
Pro Asn Gly Ile Gly Ala Leu Lys Leu Arg Xaa Met Glu His Tyr Phe
  1             5             10             15
```

Ser Gln Gly Leu Ser Val Gln Leu Phe Asn Asp Gly  
20 25

```
<210> 254
<211> 28
```

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 254

Ser Lys Gly Lys Leu Asn His Leu Cys Gly Ala Asp Phe Val Lys Ser  
1 5 10 15

His Gln Lys Pro Pro Gln Gly Met Glu Ile Lys Ser  
20 25

&lt;210&gt; 255

&lt;211&gt; 28

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 255

Asn Glu Arg Cys Cys Ser Phe Asp Gly Asp Ala Asp Arg Ile Val Tyr  
1 5 10 15

Tyr Tyr His Asp Ala Asp Gly His Phe His Leu Ile  
20 25

&lt;210&gt; 256

&lt;211&gt; 28

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 256

Asp Gly Asp Lys Ile Ala Thr Leu Ile Ser Ser Phe Leu Lys Glu Leu  
1 5 10 15

Leu Val Glu Ile Gly Glu Ser Leu Asn Ile Gly Val  
20 25

&lt;210&gt; 257

&lt;211&gt; 28

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 257

Val Gln Thr Ala Tyr Ala Asn Gly Ser Ser Thr Arg Tyr Leu Glu Glu  
1 5 10 15

Val Met Lys Val Pro Val Tyr Cys Thr Lys Thr Gly  
20 25

&lt;210&gt; 258

&lt;211&gt; 28

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 258

Val Lys His Leu His His Lys Ala Gln Glu Phe Asp Ile Gly Val Tyr  
1 5 10 15

Phe Glu Ala Asn Gly His Gly Thr Ala Leu Phe Ser  
20 25

<210> 259  
<211> 28  
<212> PRT  
<213> Homo sapiens

<400> 259  
Thr Ala Val Glu Met Lys Ile Lys Gln Ser Ala Glu Gln Leu Glu Asp  
1 5 10 15

Lys Lys Arg Lys Ala Ala Lys Met Leu Glu Asn Ile  
20 25

<210> 260  
<211> 28  
<212> PRT  
<213> Homo sapiens

<400> 260  
Ile Asp Leu Phe Asn Gln Ala Ala Gly Asp Ala Ile Ser Asp Met Leu  
1 5 10 15

Val Ile Glu Ala Ile Leu Ala Leu Lys Gly Leu Thr  
20 25

<210> 261  
<211> 28  
<212> PRT  
<213> Homo sapiens

<400> 261  
Val Gln Gln Trp Asp Ala Leu Tyr Thr Asp Leu Pro Asn Arg Gln Leu  
1 5 10 15

Lys Val Gln Val Ala Asp Arg Arg Val Ile Ser Thr  
20 25

<210> 262  
<211> 28  
<212> PRT  
<213> Homo sapiens

<220>  
<221> SITE  
<222> (2)  
<223> Xaa equals any of the naturally occurring L-amino acids

<400> 262  
Thr Xaa Ala Glu Arg Gln Ala Val Thr Pro Pro Gly Leu Gln Glu Ala  
1 5 10 15

Ile Asn Asp Leu Val Lys Lys Tyr Lys Leu Ser Arg  
20 25

<210> 263  
 <211> 24  
 <212> PRT  
 <213> Homo sapiens

<400> 263  
 Ala Phe Val Arg Pro Ser Gly Thr Glu Asp Val Val Arg Val Tyr Ala  
           1                          5                          10                          15

Glu Ala Asp Ser Gln Glu Ser Ala  
                           20

<210> 264  
 <211> 26  
 <212> PRT  
 <213> Homo sapiens

<400> 264  
 Asp His Leu Ala His Glu Val Ser Leu Ala Val Phe Gln Leu Ala Gly  
           1                          5                          10                          15

Gly Ile Gly Glu Arg Pro Gln Pro Gly Phe  
                           20                          25

<210> 265  
 <211> 443  
 <212> PRT  
 <213> Homo sapiens

<400> 265  
 Gly Thr Arg Ala Ala Pro Gly Leu Gly Ala Trp Gly Arg Arg Ser Pro  
           1                          5                          10                          15

Pro Ser Phe Ser Pro Pro Arg Pro Arg Arg Pro Gly Val Met Ala Gly  
                           20                          25                          30

Leu Asn Cys Gly Val Ser Ile Ala Leu Leu Gly Val Leu Leu Leu Gly  
           35                          40                          45

Ala Ala Arg Leu Pro Arg Gly Ala Glu Ala Phe Glu Ile Ala Leu Pro  
           50                          55                          60

Arg Glu Ser Asn Ile Thr Val Leu Ile Lys Leu Gly Thr Pro Thr Leu  
           65                          70                          75                          80

Leu Ala Lys Pro Cys Tyr Ile Val Ile Ser Lys Arg His Ile Thr Met  
                           85                          90                          95

Leu Ser Ile Lys Ser Gly Glu Arg Ile Val Phe Thr Phe Ser Cys Gln  
           100                          105                          110

Ser Pro Glu Asn His Phe Val Ile Glu Ile Gln Lys Asn Ile Asp Cys  
           115                          120                          125

Met Ser Gly Pro Cys Pro Phe Gly Glu Val Gln Leu Gln Pro Ser Thr



130	135	140
Ser Leu Leu Pro Thr Leu Asn Arg Thr Phe Ile Trp Asp Val Lys Ala		
145	150	155 160
His Lys Ser Ile Gly Leu Glu Leu Gln Phe Ser Ile Pro Arg Leu Arg		
	165	170 175
Gln Ile Gly Pro Gly Glu Ser Cys Pro Asp Gly Val Thr His Ser Ile		
	180	185 190
Ser Gly Arg Ile Asp Ala Thr Val Val Arg Ile Gly Thr Phe Cys Ser		
	195	200 205
Asn Gly Thr Val Ser Arg Ile Lys Met Gln Glu Gly Val Lys Met Ala		
	210	215 220
Leu His Leu Pro Trp Phe His Pro Arg Asn Val Ser Gly Phe Ser Ile		
	225	230 235 240
Ala Asn Arg Ser Ser Ile Lys Arg Leu Cys Ile Ile Glu Ser Val Phe		
	245	250 255
Glu Gly Glu Gly Ser Ala Thr Leu Met Ser Ala Asn Tyr Pro Glu Gly		
	260	265 270
Phe Pro Glu Asp Glu Leu Met Thr Trp Gln Phe Val Val Pro Ala His		
	275	280 285
Leu Arg Ala Ser Val Ser Phe Leu Asn Phe Asn Leu Ser Asn Cys Glu		
	290	295 300
Arg Lys Glu Glu Arg Val Glu Tyr Tyr Ile Pro Gly Ser Thr Thr Asn		
	305	310 315 320
Pro Glu Val Phe Lys Leu Glu Asp Lys Gln Pro Gly Asn Met Ala Gly		
	325	330 335
Asn Phe Asn Leu Ser Leu Gln Gly Cys Asp Gln Asp Ala Gln Ser Pro		
	340	345 350
Gly Ile Leu Arg Leu Gln Phe Gln Val Leu Val Gln His Pro Gln Asn		
	355	360 365
Glu Ser Asn Lys Ile Tyr Val Val Asp Leu Ser Asn Glu Arg Ala Met		
	370	375 380
Ser Leu Thr Ile Glu Pro Arg Pro Val Lys Gln Ser Arg Lys Phe Val		
	385	390 395 400
Pro Gly Cys Phe Val Cys Leu Glu Ser Arg Thr Cys Ser Ser Asn Leu		
	405	410 415
Thr Leu Thr Ser Gly Ser Lys His Lys Ile Ser Phe Leu Cys Asp Asp		
	420	425 430
Leu Thr Arg Leu Trp Met Asn Val Glu Lys Pro		
	435	440

<210> 266  
 <211> 159  
 <212> PRT  
 <213> Homo sapiens

<400> 266  
 Phe Glu Ile Ala Leu Pro Arg Glu Ser Asn Ile Thr Val Leu Ile Lys  
 1 5 10 15  
 Leu Gly Thr Pro Thr Leu Leu Ala Lys Pro Cys Tyr Ile Val Ile Ser  
 20 25 30  
 Lys Arg His Ile Thr Met Leu Ser Ile Lys Ser Gly Glu Arg Ile Val  
 35 40 45  
 Phe Thr Phe Ser Cys Gln Ser Pro Glu Asn His Phe Val Ile Glu Ile  
 50 55 60  
 Gln Lys Asn Ile Asp Cys Met Ser Gly Pro Cys Pro Phe Gly Glu Val  
 65 70 75 80  
 Gln Leu Gln Pro Ser Thr Ser Leu Leu Pro Thr Leu Asn Arg Thr Phe  
 85 90 95  
 Ile Trp Asp Val Lys Ala His Lys Ser Ile Gly Leu Glu Leu Gln Phe  
 100 105 110  
 Ser Ile Pro Arg Leu Arg Gln Ile Gly Pro Gly Glu Ser Cys Pro Asp  
 115 120 125  
 Gly Val Thr His Ser Ile Ser Gly Arg Ile Asp Ala Thr Val Val Arg  
 130 135 140  
 Ile Gly Thr Phe Cys Ser Asn Gly Thr Val Ser Arg Ile Lys Met  
 145 150 155

<210> 267  
 <211> 9  
 <212> PRT  
 <213> Homo sapiens

<400> 267  
 Phe Val Arg Asp Pro Phe Val Arg Leu  
 1 5

<210> 268  
 <211> 13  
 <212> PRT  
 <213> Homo sapiens

<400> 268  
 Phe Leu Phe Val Arg Asp Pro Phe Val Arg Leu Ile Ser  
 1 5 10

<210> 269

<211> 15  
 <212> PRT  
 <213> Homo sapiens

<400> 269  
 Phe Leu Phe Val Arg Asp Pro Phe Val Arg Leu Ile Ser Ala Phe  
 1 5 10 15

<210> 270  
 <211> 380  
 <212> PRT  
 <213> Homo sapiens

<400> 270  
 Tyr Leu His Thr Ser Phe Ser Arg Pro His Thr Gly Pro Pro Leu Pro  
 1 5 10 15  
 Thr Pro Gly Pro Asp Arg Asp Arg Glu Leu Thr Ala Asp Ser Asp Val  
 20 25 30  
 Asp Glu Phe Leu Asp Lys Phe Leu Ser Ala Gly Val Lys Gln Ser Asp  
 35 40 45  
 Leu Pro Arg Lys Glu Thr Glu Gln Pro Pro Ala Pro Gly Ser Met Glu  
 50 55 60  
 Glu Asn Val Arg Gly Tyr Asp Trp Ser Pro Arg Asp Ala Arg Arg Ser  
 65 70 75 80  
 Pro Asp Gln Gly Arg Gln Gln Ala Glu Arg Arg Ser Val Leu Arg Gly  
 85 90 95  
 Phe Cys Ala Asn Ser Ser Leu Ala Phe Pro Thr Lys Glu Arg Ala Phe  
 100 105 110

Asp Asp Ile Pro Asn Ser Glu Leu Ser His Leu Ile Val Asp Asp Arg  
 115 120 125  
 His Gly Ala Ile Tyr Cys Tyr Val Pro Lys Val Ala Cys Thr Asn Trp  
 130 135 140  
 Lys Arg Val Met Ile Val Leu Ser Gly Ser Leu Leu His Arg Gly Ala  
 145 150 155 160  
 Pro Tyr Arg Asp Pro Leu Arg Ile Pro Arg Glu His Val His Asn Ala  
 165 170 175  
 Ser Ala His Leu Thr Phe Asn Lys Phe Trp Arg Arg Tyr Gly Lys Leu  
 180 185 190  
 Ser Arg His Leu Met Lys Val Lys Leu Lys Lys Tyr Thr Lys Phe Leu  
 195 200 205  
 Phe Val Arg Asp Pro Phe Val Arg Leu Ile Ser Ala Phe Arg Ser Lys  
 210 215 220  
 Phe Glu Leu Glu Asn Glu Glu Phe Tyr Arg Lys Phe Ala Val Pro Met  
 225 230 235 240

Leu Arg Leu Tyr Ala Asn His Thr Ser Leu Pro Ala Ser Ala Arg Glu  
                   245                                  250                                  255  
 Ala Phe Arg Ala Gly Leu Lys Val Ser Phe Ala Asn Phe Ile Gln Tyr  
                   260                                  265                                  270  
 Leu Leu Asp Pro His Thr Glu Lys Leu Ala Pro Phe Asn Glu His Trp  
                   275                                  280                                  285  
 Arg Gln Val Tyr Arg Leu Cys His Pro Cys Gln Ile Asp Tyr Asp Phe  
                   290                                  295                                  300  
 Val Gly Lys Leu Glu Thr Leu Asp Glu Asp Ala Ala Gln Leu Leu Gln  
                   305                                  310                                  315                                  320  
 Leu Leu Gln Val Asp Arg Gln Leu Arg Phe Pro Pro Ser Tyr Arg Asn  
                   325                                  330                                  335  
 Arg Thr Ala Ser Ser Trp Glu Glu Asp Trp Phe Ala Lys Ile Pro Leu  
                   340                                  345                                  350  
 Ala Trp Arg Gln Gln Leu Tyr Lys Leu Tyr Glu Ala Asp Phe Val Leu  
                   355                                  360                                  365  
 Phe Gly Tyr Pro Lys Pro Glu Asn Leu Leu Arg Asp  
                   370                                  375                                  380  
  
 <210> 271  
 <211> 274  
 <212> PRT  
 <213> Homo sapiens  
  
 <400> 271  
 Lys Leu Val Arg Leu Gln Val Pro Val Arg Asn Ser Arg Val Asp Pro  
   1                                  5                                  10                                  15  
 Arg Val Arg Ser Lys Ile Gly Ser Arg Arg Trp Met Leu Gln Leu Ile  
                   20                                  25                                  30  
 Met Gln Leu Gly Ser Val Leu Leu Thr Arg Cys Pro Phe Trp Gly Cys  
                   35                                  40                                  45  
 Phe Ser Gln Leu Met Leu Tyr Ala Glu Arg Ala Glu Ala Arg Arg Lys  
                   50                                  55                                  60  
 Pro Asp Ile Pro Val Pro Tyr Leu Tyr Phe Asp Met Gly Ala Ala Val  
                   65                                  70                                  75                                  80  
 Leu Cys Ala Ser Phe Met Ser Phe Gly Val Lys Arg Arg Trp Phe Ala  
                   85                                  90                                  95  
 Leu Gly Ala Ala Leu Gln Leu Ala Ile Ser Thr Tyr Ala Ala Tyr Ile  
                   100                                  105                                  110  
 Gly Gly Tyr Val His Tyr Gly Asp Trp Leu Lys Val Arg Met Tyr Ser  
                   115                                  120                                  125